Development of genetic engineering tools for the cyanobacterium *Synechocystis* PCC 6803 for advanced biofuel production

Lamya Adnan Al-Haj

Department of Structural and Molecular Biology

University College London

A thesis submitted for the degree of Doctor of Philosophy

March 2014

DECLARATION PAGE

I, Lamya Adnan Atif l-Haj confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Lamya Al-Haj

ACKNOWLEDGEMENT

In the name of Allah the most gracious the most merciful

I start my praise and thanks to my creator and my sustainer, Allah all mighty, who guided every step of my life and got me to where I am today. I thank Allah for blessing me with the most amazing family and wonderful friends who stood by me throughout my hardship.

First and far most and on the top my list I thank my amazing parents, my beloved mum and dad for being there for me at all times and supporting me by all possible means during my ups and downs. Surely, I could have not done this without them!

I thank my loving, caring, supportive and patient husband who stood by me all these years and pushed me to finish this long journey and put up with all my complains and demands and supported me to the very last minute. He was always by my side and I can't thank him enough!

I thank my lovely kids Talal and Ayman for being the sunshine of my life and being the driving force for me to complete my target.

I thank my supervisor Dr. Saul Purton who not only gave me good advice and guidance throughout the years but always managed to put a smile back on my face every time I cried in his office! Thanks Saul!

I thank my brothers and sisters who always believed in me, encouraged me and supported me whenever I needed them!

I want to especially thank my in laws who repeatedly crossed the oceans to come help take care of my kids while I spent long hours in the lab! My sweet cousins also had their share of babysitting so thanks to them all too and a big thank you to Mali for being an excellent nanni.

Special thanks go to Dr.Tabisam Khan for giving me his time to listen to my complaints and giving me valibale advice.

Of course, I would like to thank the Saul Purton group especially Laura for never saying no to any help request and continuously supporting me. Thanks to Henry for being an awesome math’s tutor and Tom for being the brains for new and exciting ideas, helping me in trouble shooting, giving me great advice and for always making me smile.

I thank Tanya so much for being my mentor in my first year and always believing

in me and lots of love to Joanna and Janet for always being very supportive, positive and encouraging. I thank Umaima for being like a sister to me, Alice for helping me here and there and Nanci for showing me how to use the Algeum and being my tour guide in China!

Big thank you to Umaima, Laura, Tom, Janet, Alice, Joanna, Stephi and Rosie for partial prove reading of my thesis.

During this long journey I got to meet some amazing people and had the chance to engage in some exciting collaborative work. I would like to thank, Judith from the Department of Environmental Engineering for running my GC-MC samples, Christopher Grant from the Department of Biochemical Engineering for engaging me in his research! Ebenezer from the Department of Biochemical Engineering who allowed me to use his fancy growth incubators and special thanks to Paul Hellier from the Department of Mechanical Engineering for all the engine tests which led to a fruitful publication!

Back home I want to thank all my friends for their support especially Wafa Al Lawati for her valuable advice and encouragement. Special thanks goes to the "Prosper team", “Go-point” and "Hema Energy" who kindly provided me with office space to write my thesis, supported me in technical issues (especially Mahmood) and for the continuous encouragement!

Last but not least I thank Sultan Qaboos University for the generous PhD funds.

Thanks to all of you!

Lamya Adnan Al-Haj

ABSTRACT

Cyanobacteria hold significant potential as platforms for the production of a wide variety of high-value products and biofuel molecules such as biohydrogen, isoprenoids and alkanes. Currently, the genome sequences of over 120 cyanobacterial species are publicly available, and techniques for the genetic manipulation of a few species are well established. However, more advanced metabolic engineering technologies are required for high-throughput production/evaluation of modified strains with improved biofuel characteristics. The most popular species for genetic studies is the freshwater species, *Synechocystis* sp. PCC 6803 – not least because it is naturally transformable and is a facultative phototroph. We demonstrated in chapter 3, that a deletion of an exonuclease gene in *Synechocystis* greatly increases transformation rates, probably through the promotion of non-HR events. However the mutant generated did not display random insertional mutagenesis but rather foreign DNA was being targeted to specific locations in the genome. On the other hand, targeted gene knock-outs and knock-ins are easily achieved in *Synechocystis* via homologous recombination (HR), but the creation of each transforming plasmid is rather laborious and involves several PCR and cloning steps. Chapter 4 of this thesis describes detailed study conducted to determine the minimum length of homologous DNA sequence flanking the foreign DNA that is necessary for efficient integration into the cyanobacterial genome. Our findings suggest that as little as 50 bases is sufficient for HR and open up the possibility of a quick, single step PCR strategy using long-tail primers to achieve any desirable knock-out or knock-in. However, the efficiency of the PCR needs to be optimised in order to make the technique more efficient. Chapter 5 describes genetic modification of the isoprenoid pathway for production of the novel high-value product geraniol (C10 monoterpene) and the fuel molecule farnesene (C15 sesquiterpene). Finally, chapter 6 exploits the possibility of expressing a biological plug-in for facilitated subtract delivery and product removal of hydrocarbons in *Synechocystis.*

ABBREVIATIONS

Amp^R	Ampicilin resistance
AMPS	ammonium persulphate
cDNA	complemetary deoxyribonucleic acid
CDS	protein coding sequences
Chl a	chlorophyll a
Chl b	chlorophyll b
CoA	coenzyme A
DIG	digoxigenin-dUTP
DMAPP	dimethylallyl pyrophosphate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DW	dry weight
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid (disodium salt)
FA	fatty acid
FAMEs	fatty acid methyl esters
FPP	farnesyl pyrophosphate
FS	farnesene synthase
GC	gas chromatography
GC/MS	gas chromatography–mass spectrometry
GES	geraniol synthase
GGPP	geranylgeranyl pyrophosphate
GPP	geranyl phyrophosphate
GOI	gene of interest
HPLC	high performance liquid chromatography
IAA	isoamyl alcohol
IgG	Immunoglobulin G
IPI	Isopentenylpyrophosphate isomerase
LT30	PCR products with 30 bp region of homology to gene X
Km^R	kanamycin resistance
NCBI	National Centre for Biotechnology Information
OD	optical density

ORF	open reading frame
Ori	Origion of replication
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
pLAH.A2	plasmid containing psbA2 promoter: LAH: Lamya Al-Haj
pLAH.nrsB	plasmid containing nrsB promoter. LAH: Lamya Al-Haj
Plasmid A	pUC.psbN.km^{R}.A
PSI	photosystem one
PSII	photosystem two
RBS	ribosome binding sites
rDNA	ribosomal deoxyribonucleic acid
RNA	ribonucleic acid
RNase	ribonuclease
RT-PCR	reverse transcription-polymerase chain reaction
SD	standard deviation
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
TAGs	triacylglycerols
TE	tris EDTA
TEMED	N, N, N', N'-tetramethylethylenediamine
Tris	tris (hydroxymethyl) aminomethane
UCL	University College London
USD	United States dollar
UTR	untranslated region
UV	ultraviolet
v/v	volume for volume
w/v	weight for volume
WT	wild type

Lists of Contents

Lists of Contents 8
List of Figures 12
List of Tables 15
List of Appendices 16
Chapter 1: INTRODUCTION 17
1.1 Cyanobacteria 18
1.1.1 General overview 18
1.1.2 Commercial applications of cyanobacteria 24
1.1.3 Cyanobacterial transformation 29
1.1.4 Genetic tools for cyanobacteria 30
1.2 *Synechocystis* sp. strain PCC 6803 41
1.2.1 General overview 41
1.2.2 Cell Structure & strain origins 43
1.2.3 Promoter systems 45
1.2.4 Plasmid vectors used for *Synechocystis* 48
1.2.5 Selectable markers used in *Synechocystis* 50
1.2 Biofuels 52
1.3.1 Choosing a biofuel-producing host 53
1.3.2 Types of biofuels 56
1.3.3 The four generations of biofuels 60
1.3.4 Challenges facing biofuel production 66
1.4 Isoprenoids 69
1.4.1 Isoprenoid biosynthesis 69
1.4.2 Isoprenoid biosynthesis in *Synechocystis* 71
1.4.2 Application of isoprenoids 72
1.5 Objectives of the presented work 73
Chapter 2: Materials and Methods 74
2.1 Chemicals 75
2.2 Strains employed, their growth and storage conditions 75
2.2.1 *Escherichia coli* bacterial strains 75
2.2.2 Growth and storage of *Escherichia coli* 75
2.2.3 *Synechocystis* strains and mutants created 76
2.2.4 Growth and storage of *Synechocystis* strains 76
2.2.5 Quantification of *Synechocystis* cultures 77
2.2.6 Growth tests on solid media “spot tests” 77
2.3 DNA techniques 78
2.3.1 Expression vectors 78
2.3.2 Plasmid DNA isolation 78
2.3.3 *Synechocystis* genomic DNA isolation 79
2.3.4 Preparation of competent *E. coli* 80
2.3.5 Polymerase chain reaction 80
2.3.6 PCR purification and sequencing 81
2.3.7 Oligonucleotide primers 81

2.3.8 Codon optimisation 82
2.3.9 Agarose gel electrophoresis 82
2.3.10 Recovery of DNA from agarose gels 82
2.3.11 Construction of *Synechocystis* expression vectors 82
2.4 Transformation technique 83
2.4.1 Transformation of competent *E. coli* 83
2.4.2 Transformation of *Synechocystis* 83
2.5 DNA analysis 84
2.5.1 Southern blotting 84
2.5.2 Probe labelling and detection 85
2.6 Protein analysis 85
2.6.1 Isolation of crude whole cell protein extracts of *Synechocystis* 85
2.6.2 Protein gel electrophoresis 85
2.6.3 Western blot analysis 87
2.6.4 Immuno-detection 87
2.7 *Synechocystis* toxicity tests 89
2.7.1 Testing for growth of *Synechocystis* in the presence of terpenes 89
2.7.2 Tolerance of established cultures of *Synechocystis* to terpenes 89
2.7.3 Effect of a two-phase culture system on tolerance levels 90
2.8 Biofuel and hydrocarbon analysis 90
2.8.1 Supernatant extracts from cyanobacteria cells 90
2.8.2 Cell extracts 91
2.8.3 Two-phase extraction method 92
2.8.4 Gas Chromatography analysis 94
2.9 Growth analysis 94
2.10 Cell density determination 95
Chapter 3: Attempts for the development of a simple system for random insertion mutagenesis in *Synechocystis* 6803 96
3.0 Introduction 97
3.1 Background 99
3.1.1 Current random methods of mutagenesis 99
3.2 Chapter aims 100
3.3 Results and discussion 101
3.3.1 Creation of the Δ*recJ Synechocystis* mutant 101
3.3.2 Transformation of the Δ*recJ Synechocystis* mutant 102
3.3.3 Analysis of Δ*recJ Synechocystis* transformed with the km^R cassettes 105
3.3.4 Creation of the Δ*rec*J.km *Synechocystis* mutant. 117
3.3.5 Transformation of the Δ*rec*J.km strain with the *ble* cassette 119
3.3.6 Analysis of Δ*recJ*.km *Synechocystis* transformed with the *ble* cassette. 121
3.4 Conclusion and future work 123
Chapter 4: Development of a 'one-step' PCR-based technique for insertion of foreign DNA into *Synechocystis* 6803 126
4.0 Introduction 127
4.1 Literature review 128
4.1.1 Current methods for gene replacement in *Synechocystis* 128
4.1.2 The minimum length for homologous recombination 130

4.2 Chapter aims 131
4.3 Results and Discussion 132
4.3.1 Identification of the minimum flanking length necessary for transformation 132
4.3.2 Testing the one-step PCR-based method for generating gene knockouts in *Synechocystis* 136
4.3.3 Improving HR:NHR ratio of transformants obtained by one-step PCR method. 147
4.3.4 Comparison of PCR products from the traditional and one-step PCR methods: 157
4.3.5 Utilisation of Δ*recJ*.km^R strain and new gene knockouts for one-step PCR 162
4.4 Conclusion and future work 170
Chapter 5: The potential of engineering *Synechocystis* 6803 for photosynthetic terpene production . 172
5.0 Introduction 173
5.1 Literature review 174
5.1.1 Terpene production as high value products and biofuels 174
5.1.2 Toxicity of terpenes on micro-organisms 176
5.1.3 Physical properties and engine performance of advanced biofuels 177
5.2 Chapter background and aims 179
5.3 Results and discussion 179
5.3.1 Constructing and testing the pLAH.A2 and pLAH.nrsB expression vectors 179
5.3.2 Potential for geraniol production in *Synechocystis* 189
5.3.3 Toxicity tests of five selected terpenes *on Synechocystis* 200
5.3.4 Expression of a α-farnesene synthase gene in *Synechocystis* 206
5.3.5 Optimising α-farnesene synthase expression 210
5.3.6 Farnesene production assays 215
5.3.7 Growth analysis & phenotype observation 227
5.4 Conclusion and future work 231
Chapter 6: The expression of AlkL in *Synechocystis* to facilitate substrate delivery and product removal of hydrocarbons 234
6.0 Introduction 235
6.1 Literature review 236
6.1.1. Alkanes 236
6.1.2 The AlkL transporter 238
6.2 Aims 242
6.3 Results and discussion 243
6.3.1 Expression of *alkL* gene in *Synechocystis* under different promoters 243
6.3.2 Phenotype analysis of the AlkL expressing strain 248
6.3.3 GC analysis of *alkL* expressing strains 252
6.4 Conclusion and future work 254
Chapter 7: GENERAL DISCUSSION 256
7.1 Main Findings 257
7.1.1 Attempts for the development of a simple system for random insertion mutagenesis in *Synechocystis* 6803 257
7.1.2 Development of a 'one-step' PCR-based technique for insertion of foreign DNA into *Synechocystis* 6803 260
7.1.3 The potential of engineering *Synechocystis* 6803 for photosynthetic terpene production 265

7.1.4 The expression of AlkL in *Synechocystis* to facilitate substrate delivery and product removal of hydrocarbons 270
7.2 Conclusions, challenges and perspectives 273
References 275
Appendices 295
Publications 333

List of Figures

Figure 1.1: A picture display of variouse cyanobacterial species. 20
Figure 1.2: The stucture of phycobilisomes and the absorption spectra of various pigments. 21
Figure 1.3: Examples of cyanobacterial symbiotic associations. 23
Figure 1.4 Central metabolic pathways and products from *Synechocystis* 6803. 27
Figure 1.5: Commercial opportunities of cyanobacteria for production of biofuels and co-products. Modified from (Parmar et al., 2011). 28
Figure 1.6: A schematic representation of a typical BioBrick plasmid. 32
Figure 1.7: An illlistration of the in cis and in trans gene modification methods. 35
Figure 1.8: Methods used for targeted gene mutagenesis. 37
Figure 1.9: The *sacB* method for creating markless mutants. 40
Figure 1.10: *Synechocystis* sp. strain PCC 6803. 41
Figure 1.11: The envelope layers of the *Synechosystis* cell. 44
Figure 1.12: Strain history of *Synechosystis* sp. PCC 6803 (Ikeuchi and Tabata, 2001). 45
Figure 1.13: An illustration of the plasmids used for cyanobacterial transformation 50
Figure 1.14: Representative structures of the different fuel types 58
Figure 1.15: Comparison fo the typical bioprocess steps needed for 4 generations of biofuels production 65
Figure 1.16: Examples of different photobioreactor designs 69
Figure 1.17: Isoprenoid biosynthesis from the MEV and MEP pathways. 71
Figure 2.1: Schematic representation of extraction methods 93
Figure 3.1: Illustration of the proposed method of random km^R cassette integration into a *recJ*- strain. 98
Figure 3.2: PCR confirmation for the isolation of a Δ*recJ Synechocystis* mutant 102
Figure 3.3: Spot tests of WT, a kanamycin-resistant transformant line (*psbN*.km) and a representative *recJ::ble* transformant line. 102
Figure 3.4: Transformation of WT and Δ*recJ* strains with the kanamycin-resistance cassette 104
Figure 3.5: PCR confirmation of km^R integration in the Δ*recJ* strain. 105
Figure 3.6: Southern blot analysis of Δ*Rec*J.ble transformants. 107
Figure 3.7: Southern blot to test for genomic integration of km^R in the ∆*recJ* strain 108
Figure 3.8: Southern blot of genomic DNA digested with eight enzymes and probed with a km^R probe 110
Figure 3.9: km^R cassette extracted as *Xba*I and *EcoR*V fragments and cloned in p.JET 110
Figure 3.10: Sequance alignment and schematic representation of the Δ*recJ*.ble.km(SalI) transformant. 112
Figure 3.11: PCR results confirming the presence of km^R cassette in the *spe*A/*psbA*II locus. 113
Figure 3.12: Illustration of the insertion of the km^R marker in the *psbAII* locus. 115
Figure 3.13: The annotated sequence of the Δ*rec*J.ble.km.*Hinc*II transformant. 117
Figuer 3.14: The construction of the pRecJ::Km plasmid 118
Figure 3.15: PCR confirmation for the isolation of Δ*recJ*.km strain. 119
Figure 3.16: Transformants obtained from transformation of *ble* cassette in Δ*recJ*.km strain 120
Figure 3.17: Zeocin-resistant colonies obtained from transformation of the *ble* cassette in the WT 6803 strain. 121
Figure 3.18: Southern blot analysis of three sets of transformants (T1-T3) from the Δ*recJ*.km strain transformed with the *ble* cassette. 122
Figure 3.19: PCR analysis to investigate the itegration pattern of Δ*recJ*.km transformed with ble cassette. 123
Figure 4.1: A schematic representation of double homologous recombination in *Synechocystis*. 127

Figure 4.2: Comparison between the traditional and one-step PCR methods in creating gene knockouts in *Synechocystis*. 130
Figure 4.3: A schematic representation of the PCR products with different flank sizes. 133
Figure 4.4: PCR amplification of *psbN*::km^R knockouts with different flanking sizes. 133
Figure 4.5: Transformants generated using the traditional method. 135
Figure 4.6: PCR results confirming isolation of *psbN*::km^R knockouts using the different PCR products. 136
Figure 4.7: PCR amplification of *psbN*::km^R knockouts using one-step PCR method. 138
Figure 4.8: Transformants obtained by one-step PCR using short flank PCR products. 140
Figure 4.9: PCR results of transformants obtained using products from a one-step PCR. 141
Figure 4.10: Analysis of transformants obtained by one-step PCR with short flanks. 143
Figure 4.11: Transformants obtained with the traditional and one-step PCR methods. 145
Figure 4.12: Spot test showing tolerance of *Synechocystis psbN* knockouts to kanamycin. 145
Figure 4.13: Possible integration patterns of PCR products with short flanks obtained from the traditional and one-step PCR methods. 146
Figure 4.14: PCR results of *Synechosystis* transformed with single stranded DNA. 148
Figure 4.15: PCR results of *Synechocystis* transformed with single stranded DNA with short flanking sequences. 149
Figure 4.16: Transformants & PCR results of transformants obtained by varying DNA and cell concentration. 151
Figure 4.17: PCR analysis of transformants in Δ*recJ*.km recipient strains. 153
Figure 4.18: Applying the one-step PCR to knock out three genes using Δ*recJ*.km as recipient strain. 156
Figure 4.19: Testing the integration of PCR products of three genes with short flanks in the *speA/psbA2* area. 157
Figure 4.20: Illustrations of PCR products obtained from traditional and one-step PCR methods. 160
Figure 4.21: Illustration of *DpnI* digestion of km^R cassette in pUC.psbN::km^R plasmid. 160
Figure 4.22: Sequence results of PCR products obtained by one-step PCR and traditional methods. 161
Figure 4.23: PCR products of *psbN*, *psb28.2* and *psbA2* gene knock using *ble* marker gene. 163
Figure 4.24: Transformants of *psbN*, *psb28*.2 and *psbA*2 gene knock-outs in Δ*recJ*.km^R recipent strain. 164
Figure 4.25: PCR analysis of transformants of three gene knockouts in Δ*recJ*.km strain. 166
Figure 4.26: Transformation of WT and Δ*recJ*.km^R recipents strains with three gene knockouts with 30 bp flanks from one-step PCR. 167
Figure 4.28: PCR result of large chromosomal deletion using one-step PCR. 170
Figure 5.1: The isoprenoid pathway in *Synechosystis* 175
Figure 5.2: pLAH.A2 expression vector including the psbA2 light regulated promoter. 181
Figure 5.3: pLAH.nrsB expression vector including the nrsB inducible promoter. 183
Figure 5.4: PCR results confirming the integration of the ble marker gene in the psbA2 locus. 185
Figure 5.5: Testing the functunality of pLAH.A2 and pLAH.nrsB expression vectrors. 186
Figure 5.6: Spot tests showing the expression of active ble protein. 188
Figure 5.7: Co-inoculation of *Synechocystis* with increasing amounts of geraniol. 190
Figure 5.8: Tolerance of established cultures of *Synechocystis* to geraniol. 190
Figure 5.9: Cofirming cell death in treated cultures of *Synechosystis*. 191
Figure 5.10: Two-phase culture system used to improve tolerance towards geraniol. 193
Figure 5.11: *GES* gene isolated from pJET as a *Nde*I and *Bam*HI fragment. 194
Figure 5.12: A schematic representation showing cloning of *GES* gene under the *psbA2* and *nrsB* promoters. 194
Figure 5.13: Test digests on the pLAH.A2.GES plasmid. 195

Figure 5.14: Test digests on putative pLAH.A2.GES.F and pLAH.A2.GES.D plasmids..... 197
Figure 5.15: PCR results of *Synechocystis* transformed with full and deleted *GES*..... 199
Figure 5.16: *Synechocystis* transformed with a partially-deleted version of *GES*. 200
Figure 5.17: Testing for *Synechocystis* growth in the presence of five selected terpenes. 202
Figure 5.18: The tolerance of *Synechocystis* to increasing amounts of terpenes (24 hours). 203
Figure 5.19: The tolerance of *Synechocystis* to increasing amounts of teperens (four days post addition). 204
Figure 5.20: Effect of adding n-dodecane on survival of *Synechocystis* in the presence of increasing amounts of farnesene. 205
Figure 5.21: Effect of n-dodecane on *Synechocystis* survival in the presence of highly toxic terpenes. 206
Figure 5.22: Test digest confirming insertion of FS into pLAH.A2 and pLAH.nrsB expression vectors. 207
Figure 5.23: *Synechocystis* transformants obtained using FS in pLAH.A2.FS and pLAH.nrsB.FS plasmids. 208
Figure 5.24: Western blot analysis confirms expression of the α-farnesene synthase gene in *Synechocystis* transgenic lines. 209
Figure 5.26: Effect of high light treatment on FS expression levels. 211
Figure 5.27: Levels of α-farnesene synthase at different concentrations of the inducer. 213
Figure 5.28: Effect of induction duration on the levels of α-farnesene synthase. 214
Figure 5.30: Standards of farnesene in hexane at concentrations of 1 ppm, 50 ppm, and 100 ppm. 218
Figure 5.31: GC-MS results for samples containing FS extracted in three different solvents. 219
Figure 5.32: GC-MS results for standards of farnesene in methanol 220
Figure 5.33: Testing the presene of farnesene in samples with increased biomass. 221
Figure 5.34: GC-MS results of standards of farnesene in decane. 223
Figure 5.35: GC-MS results of negative controls prepared in decane. 224
Figure 5.36: GC-MS results of samples spiked with farnesene and extracted in decane. 225
Figure 5.37: GC-MS results of decane supernatant and cell extracts of *Synechocystis* strains..... 226
Figure 5.38: Determining the log phase of α-farnesene synthase expressing strain at OD_{525}. 227
Figure 5.39: Growth of FS expressing strain and negative control pre and post nickel induction. 229
Figure 5.40: Growth of FS expressing strain versus negative control pre and post Ni^{2+} induction..... 230
Figure 6.1: Schematic overview of alkane (alkene), fatty acid and main competing metabolic pathways in *Synechocystis*. 237
Figure 6.2: Structural comparison between the homology model of AlkL and the X-ray crystal structure for *E.coli* OmpW. 240
Figure 6.3: A schemetic representation of the proposed mechanism for alkane import. 241
Figure 6.4: A schemetic representation of the proposed mechanism for alkane export. 242
Figure 6.5: The amplification of *alkL* gene with an HA tag. 243
Figure 6.6: Transformation of *Synechocystis* with pLAH.nrsB.alkL construct and PCR confirmation. 245
Figure 6.7: The induced expression of *alkL* gene product in *Synechocystis.* 246
Figure 6.8: The titrated expression of the alkL gene in *Synechocystis.* 247
Figure 6.9: Quantification of the *alkL* gene product. 248
Figure 6.10: Phenotype analysis of the AlkL expressing strain. 250
Figure 6.11: Effect on the dry cell weight of expressing *alkL* in a single phase. 252
Figure 6.12: GC-FID analysis of Heptadecane in *alkL* expressing strains. 254
Figure 7.1: Schemetic representation of the Δcox::sacBnptI, the Δcox::sacBnptIΔrecJ::ble and the Δcox::aadAΔrecJ::ble strains. 264
Figure 7.2 : The expression of various proteins in the pLAH.A2 and pLAH.nrsB expression vectors 269

List of Tables

Table 1.1: Biofuels produced in genetically engineered cyanobacteria..... 26
Table 1.2: Cyanobacterial strains and methods used for their transformation. 30
Table 1.3: Growth modes of *Synechocystis*. Adopted from (Vermaas, 1996). 42
Table 1.4: List of promoters used to drive expression of genes in *Synechocystis*. 46
Table 1.5: Selectable markers (antibiotics resistance markers) suitable for use in *Synechocystis*. 52
Table 1.6: Biofuel alternatives to the conventional transport fuels. 58
Table 2.1: Recipe of BG11 medium for *Synechocystis* 77
Table 2.2: Plasmids, their source and usage. 78
Table 2.3: Laemmli gel recipe for SDS-PAGE (Laemmli, 1970). 86
Table 2.4: Details of primary and secondary antibodies. 88
Table 4.1: Transformation results of LT50.*psbN*::km^R PCR products in Δ*recJ*.ble and WT recipient cells. . 153
Table 5.1: Ranking of 12 selected terpenes from best to worst performance in diesel engine. 178
Table 6.1: The set up used for phenotype analysis of AlkL expressing strain..... 249
Table 7.1: Expected antibiotic resistance patterns of the three strains. 263

List of Appendices

Appendix 1: Plasmids utilised in this work. 296
Appendix 2: Recipe for stock solutions preparation of BG11 medium 301
Appendix 3: List of primers 302
Appendix 4: CDS of GES 306
Appendix 5: Full sequence of RT17 (CT-TOPO) plasmid 307
Appendix 6: Annotated sequence of the RecJ.km.HincII transformant vrs WT6803 sequence. 308
Appendix 7: Sequence results of LT50.psbA2.km^R PCR product 311
Appendix 8: Operon responsible for phycobilisome expression 312
Appendix 10: Sequenced analysis of pLAH.A2.GES.F expression vector 315
Appendix 11: Condon optimised α-farnesene synthase gene and its corresponding protein sequence. 316
Appendix 12: GC-MS results of Farnesene in n-dodecane. 319
Appendix 13: pLAH.nrsB.GES.HA sequence results 322
Appendix 14: Full sequence and map of the pJ811 plasmid including the alkL gene (red) 325
Appendix 15: Sequence results of the *alkL* gene with an HA tag 327
Appendix 16: List of websites used in this work. 328
Appendix 17: Sequence analysis of α–farnesene synthase in the pLAH.A2 and pLAH.nrsB plasmids. 329

Chapter 1: INTRODUCTION

1.1 Cyanobacteria

Cyanobacteria (traditionally known as blue-green algae) represent an extremely diverse, yet highly specialised group of prokaryotic organisms that live in diverse environmental habitats (Hu et al., 2008). Cyanobacteria, which are considered ancient relatives of chloroplasts (McFadden, 1999), are the only prokaryotic organisms capable of carrying out both plant-like oxygenic photosynthesis (Ikeuchi and Tabata, 2001) and hydrogen production (Parmar et al., 2011). The photosynthetic apparatus involves the cooperation of two photosystems; photosystem I (PSI) and photosystem II (PSII) (Chauvat et al., 1989, Golden et al., 1987). Furthermore, they are the only prokaryotic organisms exhibiting a circadian clock mechanism (Kunert et al., 2000). Cyanobacteria inhabited earth for billions of years during the precambrian era (Mazel et al., 1990) making them one of the oldest organisms (Knoop et al., 2010). It has been suggested that they played an important role in preserving organic matter found in ancient sediments (Tan et al., 2011, Golden et al., 1987) and are believed to be responsible for the transformation of the earth's atmosphere from anoxic to oxic state (Koksharova, 2002, Schwartz, 1975). Cyanobacteria, being amongst the first colonizers of land, provided both physical and chemical substrate for the later growth of plants (Koksharova, 2002). Therefore, it comes as no surprise that they have played a key role in both evolution and shaping of the earth's ecology (Vermaas, 2007).

1.1.1 General overview

Cyanobacteria are very interesting organisms as they display diversity in their metabolism and structure, morphology and habitats (Figure 1.1) (Deng and Coleman, 1999). Although classified as photosynthetic bacteria, they appear to be the phylogenetic link between bacteria and algae (Stanier et al., 1971). This is because, despite their cellular organization being prokaryotic, their photosynthetic physiology is very similar to that of the chloroplasts of algae and higher plants (Labarre et al., 1989, Koksharova, 2002). Despite the fact that cyanobacteria form a phylogenetically monophyletic group of Gram-negative bacteria (Porter, 1986), they are still referred to as microalgae today (Hu et al., 2008). The distinct blue-green colour of cyanobacterial cultures (Figure 1.2.A) is attributed to the combination of the green chlorophyll pigment (chlorophyll *a*) and the blue pigment phycocyanin (Figure 1.2.B and C) (Whitton and Potts, 2000). Carotenoids are

also amongst other pigments found in cyanobacteria (Figure 1.2.C) (Stanier et al., 1971). There are over 4000 distinct species of cyanobacteria (Dismukes et al., 2008) and unlike most other prokaryotes (bacteria or archaea), they are morphologically recognisable (even at the species level) before collection (Castenholz, 1988) because of their conspicuous traits. Despite the fact that cyanobacteria lack flagella, some species seem to show a gliding type of movement when in contact with a solid surface, while a few are known to swim in liquid (Ikeuchi and Tabata, 2001, Koksharova, 2002). Moreover, various cyanobacteria are known to respond to light, a process known as phototaxis, resulting in preferential localization under certain light conditions. However, the exact mechanisms of how positive (towards light) and negative (away from light) phototaxis is controlled, and the overall motility of this cyanobacterium remain unclear (Ikeuchi and Tabata, 2001).

Figure removed for copyright reasons

Figure 1.1: A picture display of variouse cyanobacterial species.

Cyanobacterial species including: (A) *Spirulina* and *Lyngbya,* (B) *Nostoc (*arrows pointing at heterocyst's), (C) *Lyngbya,* (D) *Merismopedia*, (E) *Spirulina*, (F) *Synechocystis.* Imagines were taken from the internet as detailed in Appendix 16.

Figure removed for copyright reasons

Figure 1.2: The stucture of phycobilisomes and the absorption spectra of various pigments.

(A) A culture of *Synechocystis* 6803 (B) The structure of the phycobilisome chromophore, (C) The adsorption spectra of various pigments present in cyanobacteria (Chlorophyl a and b, carotenoids and phycobilins). (Image A is from my personal collection while B and C were taken from the internet as detailed in Appendix 16.

Cyanobacteria can be found almost everywhere on the surface of earth since they can grow and survive in numerous environmental conditions (Vermaas, 2007). They are particularly common in aqueous environments including freshwater, brackish, marine and hypersaline (Hu et al., 2008), non-acidic hot springs and in soils, rocks and deserts (Ikeuchi and Tabata, 2001, Koksharova, 2002). In temperate regions, cyanobacteria are especially common in calcareous and alkaline soil. They can also be found in extreme environments, such as acidic soils (*e.g. Anabaena oscillarioides*), hot springs (*e.g. Synechococcus lividus*) or moist terrestrial environments of polar regions (*e.g. Nostoc fuscescens*) (Castenholz, 1988).

Cyanobacteria play a vital role in diverse ecological systems as they often form symbiotic relationships with many organisms including plants, algae, fungi, bryophytes, ferns, cycads, and invertebrates (corals) (Schwartz, 1975, Adams and Duggan, 2008, Adams, 2002). Examples of such symbiotic associations are presented in Figure 1.3 and are shown to play a vital role in the recycling of carbon and nitrogen in the atmosphere (Ruffing, 2011).

Cyanobacteria are generally grouped into unicellular, colonial, pseudoparenchymatous, unbranched filamentous, heterocystous, and

heterotrichous-heterocystous forms (Desikachary, 1959). Their cell walls are very thin, but are frequently surrounded by a broad mucilaginous sheath that may be coloured (Prosperi, 2000). Cyanobacterial genomes range from 1.7 Mb for *Prochlorococcus marinus* MED4 which is a marine organism believed to be the smallest free-living bacterial genome sequenced thus far, to 9.2 Mb for the largest genome that belongs to *Nostoc punctiforme* ATCC 29133 (Vermaas, 2007).

There are several properties in cyanobacteria that make them target organisms to study many important processes. For example, cyanobacteria being photosynthetic, grow and differentiate when provided with simply sunlight, air, and water (Koksharova, 2002). Furthermore, being prokaryotes they are in principle, amenable to genetic manipulation by the techniques already developed for the intensively studied heterotrophic bacteria such as *E. coli* (Porter, 1986). Moreover, the availability of the full annotated genomic sequence of many cyanobacteria makes their genetic manipulation even more attractive (Vermaas, 2007). For the reasons mentioned above, they have been widely used to study, among other topics, higher plant processes like photosynthesis and its genetic control, nitrogen fixation, gene regulation and metabolism and heterocyst differentiation (Koksharova, 2002, Dzelzkalns and Bogorad, 1986, Min and Sherman, 2010a).

Figure removed for copyright reasons

Figure 1.3: Examples of cyanobacterial symbiotic associations.

(A) Symbiotic association between a dinoflagellate (green) with two morophotypes (arrows) of yellow fluorescent cyanobacteria inside the girdle of the dinoflagellate (Image taken by RS Foster in the subtropical Atlantic). (B) Symbiotic association between the aquatic fern *Azolla* and the nitrogen fixing cyanobacterium *Anabaena*, (C) Symbiotic association between the nitrogen fixing cyanobacterium *Nostoc* with the fungi *Geosiphon pyriformis* showing *Nostoc* filament (red) and the extracellular polysaccharides of *Nostoc* and the outer layer of the fungal cell wall labelled by the fluorescence-coupled lectin ConA (green).Images were taken from the internet as detailed in Appendix 16.

1.1.2 Commercial applications of cyanobacteria

Microorganisms, in general, have long been exploited for their natural products, some of which have been used as fuels, commodity chemicals, specialty chemicals, polymers, and drugs (Mukhopadhyay et al., 2008). Cyanobacterial strains in particular, have been used for many applications as they can produce a range of high-value chemicals and high-energy fuels from sustainable resources (Tan et al., 2011) some of which will be covered in this section.

Although the use of cyanobacteria as food supplement is currently restricted to a few species (*Spirulina (Arthospira)*, *Nostoc* and *Aphanizomenon*) (Spolaore et al., 2006), cyanobacteria have commonly been used as a food supplement since early history. In China for example*, Nostoc flagelliforme* is considered a delicacy (Gao, 1998) while other species are eaten in India and the Philippines (Tiwari, 1978). Species of *Spirulina* are used as a rich source of protein and vitamins (protein accounting for ~ 70% of the dry cell weight) and the protein is used for both humans and animals (Ciferri and Tiboni, 1985, Parmar et al., 2011). Cyanobacterial pigments are also used as a natural colourant in food (Kay and Barton, 1991). Recently, unicellular cyanobacteria have shown great potential for the production of valuable products (Knoop et al., 2010). For example, since cyanobacterial extracts have been shown to display antioxidant properties (Miller et al., 1996), they have been used in the prevention of chronic diseases (Rao and Agarwal, 1999) and in drug discovery (Tan, 2007, Sielaff et al., 2006). This is attributed to the ability of some strains to synthesise secondary metabolites with reported therapeutic effects (Belay et al., 1993).

Biomolecules such as phycobiliproteins, phcoerythrin, phycocyanin and allophycocyanin are naturally produced by cyanobacteria and have many commercial applications (Parmar et al., 2011). For example, phycobiliproteins are commonly used in fluorescent tagging applications in research (Ruffing, 2011) in conjugation with monoclonal and polyclonal antibodies to make fluorescent antibody reagents that are used for cell sorting (Koksharova, 2002). While the other natural products are used as food dyes or pigments in cosmetics (Parmar et al., 2011). Since cyanobacteria mostly thrive in aquatic environments, they are excellent vectors for the delivery of mosquitocidal toxins (Ruffing, 2011) while some strains are commonly used as natural fertilizers since they can fix

atmospheric nitrogen (Parmar et al., 2011). Other environmental applications include the use of bioluminescent cyanobacterial reporter strains. Such strains have been used as biosensors for the detection of nickel, cobalt and zinc in the environment (Loredana et al., 2008). Furthermore, the ability of some strains to degrade aromatic hydrocarbons and xenobiotics (due to associated symbiotic bacteria) have allowed for their successful applications in bioremediation (Koksharova, 2002).

Another important commercial interest is the synthesis of carotenoids, which can be used in industry as colouring agents for food, pharmaceuticals, cosmetics and in animal feed (Lagarde et al., 2000). On the other hand, genetically engineered cyanobacteria have been utilised to produce chemical products such as fish growth hormones (Ruffing, 2011), polyhydroxyalkanoates (PHA's), a class of biodegradable polyesters with many industrial applications (Koksharova, 2002, Ducat et al., 2011b) and polyunsaturated fatty acids (PUFA) that are used as both human food and animal feed for their known health promoting properties (Radmer and Parker, 1994). The use of genetically modified cyanobacteria has increased the range of products that can potentially be produced. For example, *Synechococcus* has been recently genetically engineered for the heterologous expression of a single gene (*efe*) from *Pseudomonas syringae* that lead to the successful synthesis of ethylene (Sakai et al., 1997), one of the most important building-blocks in synthetic chemical industry (Wang et al., 2012). Furthermore, acetone, a solvent and also an important precursor for industrial chemicals was recently produced by modified *Synechocystis* with a reported yield of 36.0 mg/L (Zhou et al., 2012).

On the other hand, many reports exist for the use of genetically modified (GM) cyanobacteria in the production of biofuels (Lindberg et al., 2010, Atsumi et al., 2009). Table 1.1 contains some examples from work done by various researchers of such strains and the biofuels they have been engineered to produce while Figure 1.4 shows the central metabolic pathways in *Synechoystis* that have been engineered to produce the biofuels listed in Table 1.1 and variouse other products of interest. Recently, a few companies, such as Algenol Biofuels and Joule Unlimited, have been set up with the focus on using GM cyanobacteria for large-scale biofuel production. The fact that these compounds can be produced as finished fuels that do not require additional refining or

chemical modification for use in an engine (Rude and Schirmer, 2009) and that their composition can be tailored by genetic manipulation makes microbial-produced fuel molecules excellent options as future fungible fuels (Golden et al., 1987). In addition to the applications mentioned above, cyanobacteria are being optimised to overproduce numerous products of value (Liu and Curtiss, 2009). Figure 1.5 summarises the many commercial opportunities where cyanobacteria can be utilised. It appears that with the recent hyperbole on engineered cyanobacteria, it is possible that they will become the *E. coli* of the next industrial generation (Ruffing, 2011).

Table 1.1: Biofuels produced in genetically engineered cyanobacteria.

Fuel molecule	Species engineered	No. of genes	ref
Ethanol (C_2H_5OH)	• *Synechocystis* sp PCC6803 • *Synechococcus elongatus* sp PCC7942 • *Anabaena* sp PCC7120	2 2	(Dexter and Fu, 2009) (Deng and Coleman, 1999) Algenol Biofuels 2010
1-Butanol (C_4H_9OH)	• *Synechococcus elongatus* sp PCC7942	5	(Lan and Liao, 2011)
Isobutanol (C_4H_9OH)	• *Synechococcus elongates* sp PCC7942 • *Synechocystis* sp PCC6803	5 2	(Atsumi et al., 2009) (Varman et al., 2013)
Ethylene (C_2H_4)	• *Synechococcus elongatus* sp PCC7942	1	(Sakai et al., 1997)
Isobutyraldehyde (C_4H_8O)	• *Synechococcus elongatus* sp PCC7942	5	(Atsumi et al., 2009)
Isoprene (C_5H_8)	• *Synechocystis* sp PCC6803	1	(Lindberg et al., 2010)
Free fatty acids (C_{10}–C_{18})	• *Synechocystis* sp PCC6803	5	(Liu et al., 2011)
Fatty alcohols (C_{16}/C_{18})	• *Synechocystis* sp PCC6803	1	(Tan et al., 2011)
n-alkanes (C_nH_{2n})	• *Synechococcus* sp PCC7002 • *Thermosynechococcus elongatus* sp BP-1	2	(Tan et al., 2011)

	• *Synechocystis* sp PCC6803		
Hydrogen (H_2)	• *Synechococcus elongates* sp PCC7942	1	(Ducat et al., 2011a)

Figure removed for copyright reasons

Figure 1.4 Central metabolic pathways and products from *Synechocystis* 6803.

The functions of several pathways (marked as dot-lines) are still not verified. **Key**: **3PG**: 3-phosphoglycerate; **3HB**: 3-hydroxybutyrate; **AKG**: α-ketoglutarate; **CIT**: citrate; **F6P**: fructose 6-phosphate; **G6P**: glucose 6-phosphate; **GAP**: glyceraldehyde 3-phosphate; **GLX**: glyoxylate; **GLY**: glycine; **ICT**: isocitrate; **MAL**: malate; **Mal-CoA**: Malonyl-CoA; **OAA**: oxoacetate; **PEP**: phosphoenolpyruvate; **PHB**: polyhydroxybutyrate; **PSI&PSII**: photosystem I & photosystem II; **PYR**: pyruvate; **Ru5P**: ribulose-5-phosphate; **RuBP**: ribulose-1,5-diphosphate; **SER**: serine; **SUC:** succinate; **SucCoA**: succinyl-CoA; **SSA**: succinic semialdehyde. Adopted from (Yu et al., 2013).

Figure removed for copyright reasons

Figure 1.5: Commercial opportunities of cyanobacteria for production of biofuels and co-products. Modified from (Parmar et al., 2011).

Key: BTL: bipolar tetraether lipids, CH_4: methane, O_2: oxygen, CO_2: carbon dioxide, H_2O: water, H_2: hydrogen, e-: electrons.

1.1.3 Cyanobacterial transformation

Several cyanobacterial strains are capable of being transformed with exogenously added DNA (Porter, 1986) or can accept mobilised plasmid DNA by conjugation from *E. coli* (Elhai and Wolk, 1988). Natural transformation, which is the transfer of free DNA into cells, was first described for *Synechococcus* sp. strain PCC7943 (*Anacystis nidulans* 602) by Shestakov and Khuyen (Shestakov and Khyen, 1970) and is still considered the primary means for gene transfer in unicellular cyanobacteria today (Golden et al., 1987). *Synechococcus* species along with *Synechocystis* species are among the few species that are known to be naturally transformable by exogenous DNA (Porter, 1986).

Since the discovery of cyanobacterial transformation in the 1970's, many techniques such as electroporation and ultrasonic treatment have been developed and applied to several cyanobacterial strains. Examples of strains transformed by different methods can be found in Table 1.2. It was noted that, in *Synechocystis*, almost all the exogenous DNA is integrated into the genome following natural transformation by means of a well-established double homologous recombination system (Zang et al., 2007, Kufryk et al., 2002) and cyanobacteria that use systems with homologous recombination mechanisms allow for more effective gene modifications to wild type genes (Vermaas, 1998). This is true since gene replacement depends on homologous recombination in order to replace a section of the recipient chromosome with a donor DNA that contains a selectable marker (Figure 1.7.A) (Williams, 1988). Williams and Szalay. (1983), were the first to demonstrate homologous recombination in cyanobacteria as a means of transformation, while Golden and Wiest. (1988), were the first to achieve homologous recombination in *Anabaena*.

Transformation of cyanobacteria has been thoroughly reviewed (Porter, 1986, Shestakov and Khyen, 1970, Eaton-Rye, 2004, Koksharova, 2002) and the methodology has been well described (Porter, 1988, Golden and Sherman, 1984). Nonetheless, very little is known on the mechanism by which DNA binds to the cell, gets processed and subsequently transported across the membrane during transformation (Kufryk et al., 2002, Vermaas, 1996, Barten and Lill, 1995). However, it is thought that type IV pili are involved in the uptake process (Ikeuchi and Tabata, 2001, Eaton-Rye, 2004). All of the cyanobacteria that have been

shown to be transformable are unicellular, non-heterocyst forming strains that have been classified as *Synechococcus* or *Synechocystis* (Porter, 1986). Furthermore, out of the approximately 300 cyanobacterial species registered in culture collections, less than ten have been reported to be naturally competent (de Tarsac and Houmard, 1987).

There are many factors that affect the frequency and efficiency of transformation in cyanobacteria. These include the strain used, the length, form and concentration of DNA used and also on the method used for transformation (Kufryk et al., 2002). There is no doubt that improving transformation methods in cyanobacteria will help accelerate cyanobacterial research and will aid in the development of tools for the modification of cyanobacteria.

Table 1.2: Cyanobacterial strains and methods used for their transformation.

Cyanobacterial species	Method of transformation	Reference
Anabaena sp.	▪ electroporation	(Thiel and Poo, 1989)
Synechococcus sp.	▪ electroporation ▪ natural transformation	(Matsunaga et al., 1990) (Porter, 1986)
Fremylla diplosiphon	▪ electroporation	(Chiang et al., 1992)
Plectonema boryanum	▪ electroporation	(Vachhani et al., 1993)
Synechocystis sp.	▪ electroporation ▪ ultrasonic transformation ▪ naturaltransformation	(Zang et al., 2007)

1.1.4 Genetic tools for cyanobacteria

The past two decades have shown enormous development in the field of molecular biology for cyanobacteria going from almost the absence of any knowledge of the molecular biology of cyanobacteria in the mid-seventies to the development of transformation systems in the eighties (Golden and Sherman, 1984, Golden et al., 1987) to a full-fledged unique and promising field of research in the nineties (Kay and Barton, 1991, Sakai et al., 1997, Vermaas, 1998) due to the determination of the full sequence of the cyanobacterium *Synechocystis* PCC 6803 in 1996 (Vermaas, 1996). Since then, tremendous efforts have been

invested in the development of many genetic tools for unicellular and filamentous strains of cyanobacteria and they have been thoroughly reviewed (Thiel, 2004, Elhai, 1994, Elhai and Wolk, 1988).

As mentioned in the previous section, cyanobacteria have a great potential in the production of renewable fuels, various chemicals and nutritional products. The development of synthetic biology tools can help un-lock cyanobacteria's potential for these functions. However, the development of molecular tools and engineering technologies for these microorganisms is far behind those established for *S. cerevisiae* and *E. coli* (Berla et al., 2013). Nevertheless, the recent development of genetic engineering and synthetic biology tools, has seen significant advances in strain improvement (Alper and Stephanopoulos, 2009). Synthetic biology can be defined as "the use of DNA synthesis and recombinant DNA technologies to design and construct novel functions and genetic circuits" (Alper and Stephanopoulos, 2009). According to Connor and Atsumi. (2010), synthetic biology in its core aims to (1) use standardised and well characterised building blocks "bio-Bricks", (2) hierarchical design of artificial yet nature inspired genetic circuits *in silico*, and (3) use chemical synthesis to produce novel DNA sequences to achieve a desired target (Huang et al., 2010). The steps therefore involve the creation of a "toolbox" or a "library" of biological parts that can be used separately or in a larger assembly for a higher function (Andrianantoandro et al., 2006).

The "bio-Brick" method, involves the expression of genes in microorganisms by the addition of plasmids that are constructed to include standardised genetic "parts" such as promoters, ribosome binding sites (RBS), genes, terminators and other regulatory elements that can all be swapped in and out standard plasmids as illustrated in Figure 1.6 (Berla et al., 2013, Wang et al., 2012). This method is becoming routinely used for gene expression in cyanobacteria (Huang et al., 2010, Huang and Lindblad, 2013, Landry et al., 2013).

Figure 1.6: A schematic representation of a typical BioBrick plasmid.

Key: (E: *EcoR*I and X: *Xba*I): BioBrick cloning site prefix, (S: *Spe*I and P: *Pst*I): BioBrick cloning site suffix, examples of parts include: promoters, ribosome binding sites (RBS), protein coding sequences (CDS) and terminators. The figure is based on an image taken from the internet as detailed in Appendix 16.

The assembly of the biological parts/components happens in a hierarchal manner to create a programed cell with a desired function (Andrianantoandro et al., 2006). This modern tool has become a valuable technology and has opened the doors to produce a whole new range of novel molecules in standard model organisms. For example, synthetic biology has been applied to build and introduce synthetic pathways for hydrogen production in *E. coli* (Yoshida et al., 2007) while it was employed for isoprenoid-based fuel production in various other microorganisms including cyanobacteria (Lee et al., 2008, Withers et al., 2007, Keasling and Chou, 2008).

On the other hand, metabolic engineering, which is considered as the main technology for designing biofuel producing organisms, emerged approximately 15 years ago as a distinct field that is differentiated from genetic engineering by its focus on the properties of biosynthetic and metabolic pathways in their entirety, instead of single genes and enzymes (Stephanopoulos and Sinskey, 1993). Traditional engineering approaches therefore focus on the “fine-tuning” of core components to meet strict guidelines in order to achieve a specific task (Connor and Atsumi, 2010, Keasling, 2008). Using synthetic biology for advanced metabolic engineering could help in overcoming challenges such as the

complexity of modeling the photosynthetic metabolism and poor annotations of important metabolic pathways for most cyanobacteria (Berla et al., 2013).

Model organisms: There are a number of cyanobacterial species that are suitable for genetic modification and these include the two model organisms, *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 7002. Both strains are naturally transformable and efficiently integrate foreign DNA into their genome by means of homologous recombination (Williams, 1988, Vermaas, 1998). Furthermore, *Synechocystis* has the ability to grow well in the absence photosynthesis so long as the cells are supplemented with glucose (Vermaas, 1996). The availability of fully sequenced genomes of some strains makes genetic manipulations of these organisms much more attractive as information can be extracted through functional genomics, proteome and transcriptome studies (Ikeuchi and Tabata, 2001). The wealth of information on *Synechocystis* specifically allowed for biochemical pathway modifications and engineering of this strain (Wu and Vermaas, 1995), which allowed for the development of *Synechocystis* 6803 models that analyse growth under different conditions, detect bottlenecks and gene-knockout candidates for enhanced metabolic production of products such as hydrogen and ethanol (Knoop et al., 2010), and models that represent the photosynthetic apparatus in detail and can detect alternate flow pathways of electrons (Nogales et al., 2012). The latest strain of all *Synechocystis* 6803 models is the i*Syn*731 strain, which integrates all these recent developments and adds on to them improved metabolic capability (Berla et al., 2013). For example, the strain contains as many as 322 unique reactions including those related to the heptadecane and fatty acid biosynthesis (Saha et al., 2012). Nevertheless, there is an essential need to develop a more advanced synthetic biology toolbox for cyanobacteria similar to those developed for *E. coli*.

Gene modifications: Gene modification plays a critical role in the successful development of engineered microorganisms (Ruffing, 2011) and can be performed either in cis (through chromosomal editing) or in trans (through foreign plasmid addition) where in cis gene modification is the most common approach in cyanobacterial synthetic biology (Berla et al., 2013). This is because chromosomal integration of a target gene allows for a more strict and reproducible control of the expressed gene (Ruffing, 2011). Figure 1.7 illustrates the difference between the two approaches for gene modifications. Furthermore,

advances in DNA synthesis technologies now make possible the design of bespoke coding sequences, and the use of codon optimised trans-genes has been shown to improve the expression of heterologous proteins in cyanobacterial hosts (Gustafsson et al., 2004). It is important to note that while the over-expression of genes is a useful tool to increase the overall yield of a desired product, there are several problems associated with over-expression in an engineered host and these include: the depletion of precursors or resources necessary for the host growth and production (Glick, 1995), the stress introduced in the host due to heterologous protein expression (Goff and Goldberg, 1985) and the imbalances in the total enzyme activity in the pathway that can lead to restricted carbon flux and lead to accumulation of toxic intermediate or products (Peralta-Yahya and Keasling, 2010). Nevertheless, it has been shown that by over-expressing IPP and DMAPP in *E. coli*, using terpene synthase mutants, that eight different terpenes can be produced in the engineered strains (Yoshikuni et al., 2006).

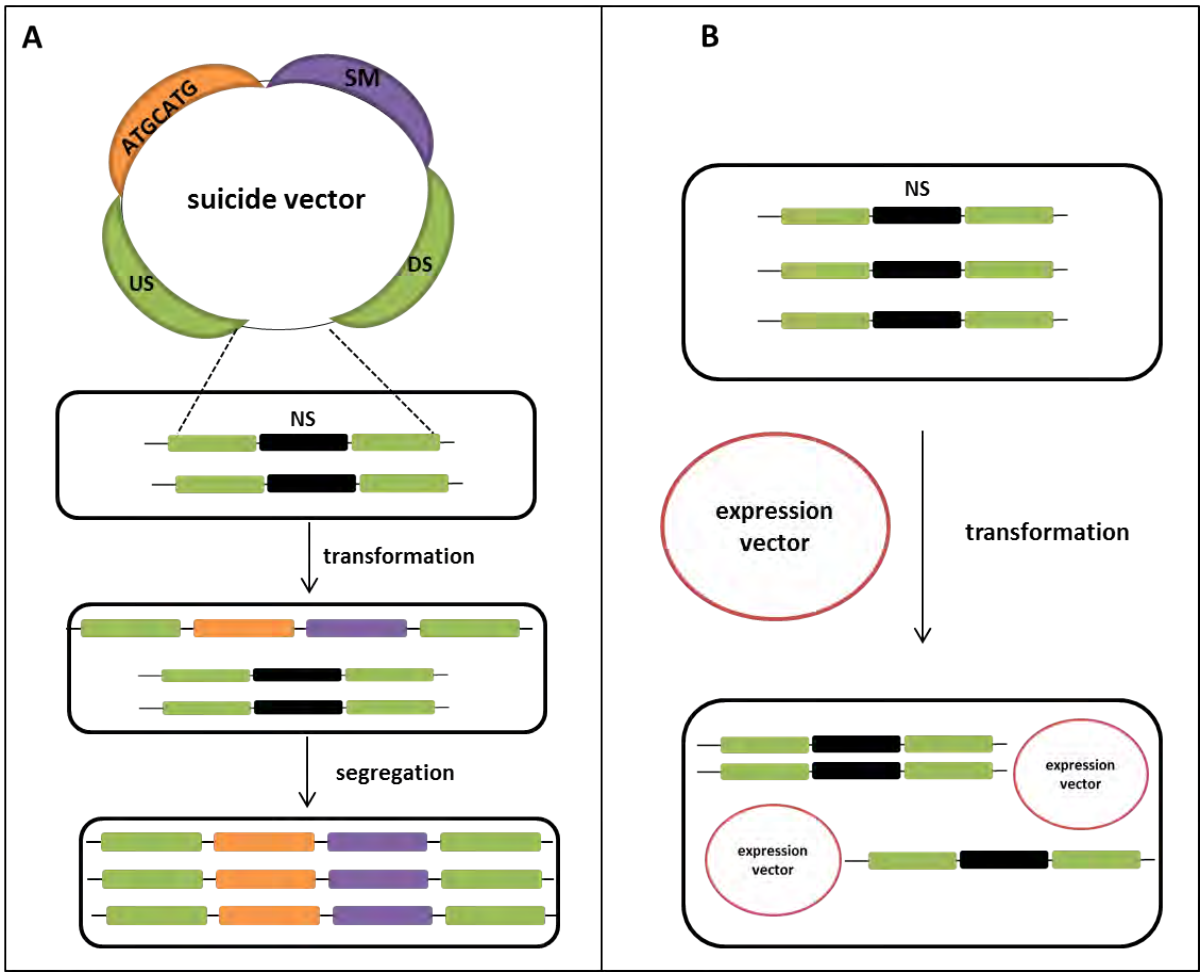


Figure 1.7: An illlistration of the in cis and in trans gene modification methods.

(A) The method for cis genetic modification employs a suicide vector that is inserted into the cyanobacterial genome at a neutral site (NS, black) with upstream (US) and downstream (DS) flanking regions in the vector (green). The insert contains an arbitrary sequence of interest (ATGCATG, orange) and a selectable marker (SM, purple). (B) Genetic modification in Trans via expression plasmids. Modified from (Berla et al., 2013).

Other advances in the field include the development of gene transfer methods for cyanobacterial species that are not naturally transformable. Furthermore, the ability to mutate or manipulate endogenous genes is necessary to complement the introduction of novel trans-genes in any metabolic engineering strategy. Gene transfer methods, in addition to natural transformation, include electroporation and conjugation systems (Berla et al., 2013, Ruffing, 2011). Electroporation involves exposing the cells to an electrical field, which results in a temporary loss of semipermeability of cell membranes, thus leading to an increased uptake by cells of DNA (Tsong, 1991), while conjugation involves the transfer of plasmid DNA from a donor to a recipient bacterium through cell-to-cell contact (Stucken et

al., 2012). Although electroporation has been successfully employed for the transformation of filamentous cyanobacteria, conjugal transfer from *Escherichia coli* is generally the standard method of choice. On the other hand, natural transformation is generally the method used for the transformation of unicellular cyanobacteria (Ravindran et al., 2006, Thiel and Poo, 1989).

While gene inactivation can be achieved by chemical and UV mutagenesis (Dzelzkalns and Bogorad, 1986), other methods are also available such as targeted gene interruption (Williams, 1988), site–directed mutagenesis (Chu et al., 1994), transposon mutagenesis (Tn5 mutagenesis) (Tandeau de Marsac et al., 1982, Berla et al., 2013), and random cartridge mutagenesis (Labarre et al., 1989, Chauvat et al., 1989). All methods are equally important and the method of choice depends largely on the particular application. However, it should be noted that the use of transposons to create mutants works poorly in cyanobacteria (Berla et al., 2013). Figure 1.8 illustrates methods used for targeted gene mutagenesis.Transformation and conjugation systems combined with the ability to clone and inactivate genes in cyanobacteria by various methods have opened the door to many advanced studies (Golden et al., 1987) and can be the first step towards many applications such as pathway engineering, overproduction of specific compounds and the biochemical analysis of protein function.

Figure removed for copyright reasons

Figure 1.8: Methods used for targeted gene mutagenesis.

(A) Targeted gene interruption by means of double homologous recombination. (B) Site-directed mutagenesis employing oligonucleotides which can include point mutations, insertions and deletions. Figure B was modified from an image taken from the internet as detailed in Appendix 16.

Pathway engineering: It is well known that the metabolic networks of any microorganism are highly complex and nonlinear. They are also tightly coordinated to carry out a large number of biochemical reactions in a highly organised manner. While the metabolic networks of bacteria have been studied for over three decades, those of cyanobacteria have only relatively recently been investigated (Lu et al., 2011). One example is the impressive efforts done by Fu. (2009), who reconstructed the genome-scale metabolic network of *Synechocystis* which included 633 genes, 704 metabolites and 831 metabolic reactions. This was done to study optimal *Synechocystis* growth, network capacity and functions as well as *in silico* predictions of different growth conditions for the model organism. Furthermore, the model was utilised for the insertion of the ethanol fermentation pathway.

It is clear from the previous example that metabolic engineering work no longer only focuses on re-wiring native metabolic pathways towards pathways of interest (Leonard et al., 2008, Zhang et al., 2008) but rather the construction of new pathways. Another example, is the enhanced production of farnesene and various long-chain alcohols (C_5-C_8) in *E. coli* by Wang et al. (2011), and Zhang et al. (2008), respectively via the re-construction of the yeast mevalonate pathway in *E. coli*. Nonetheless, it is still not clear whether metabolic pathway optimisation is more robust when it is carried out in native strains or in recombinant strains in which exogenous pathways are introduced (Alper and Stephanopoulos, 2009) and for this both methods are currently utilised.

Finally, it is important to note that in the process of optimising microbial metabolic pathways, it is essential to maintain a balanced redox state in the engineered microbes (Mukhopadhyay et al., 2008). This is to avoid the buildup of toxic byproducts that will ultimately hamper the production of a desired product. Furthermore, production potentials are limited by the low activity of pathway enzymes, an important factor that needs to be addressed when developing tools for pathway engineering. There is no doubt that adopting recently emerging tools in synthetic biology will be enormously beneficial in overcoming such challenges (Lee et al., 2008).

New tools for model organisms: Several new tools for the model organism *Synechocystis* have recently been developed. For example, a cell-viability fluorescence assay was developed that provides a quick and simple way of

detecting viable from none viable cells in a cyanobacterial culture. This method can be used to monitor the fitness of cultures grown for novel products such as biofuels (Schulze et al., 2011). Recently, the development of degradation tags used for the fine-tuning of protein levels and protein turnover times in this model organism has been achieved. Such advances aids in the generation of more complex metabolic systems that have the potential to oscillate in step with circadian rhythms in cyanobacteria (Landry et al., 2013). Furthermore, recombinant DNA technology has been employed in *Synechocystis* for the increased production of carotenoids (*e.g.* zeaxanthin) (Lagarde et al., 2000).

Another important tool in genetic engineering work that has a particularly important industrial application is the development of gene replacements methods that allow for the creation of antibiotic-free genetically modified microorganism (Lagarde et al., 2000). Such strains can be created using the toxic levanusucrase gene (*sacB)* that is isolated from *Bacillus subtilis* (Eaton-Rye, 2004, Ruffing, 2011). In this method, first the microorganism is transformed with the *sacB* gene (or any other toxic gene) linked to the selectable marker gene with flanking sequences to replace the target gene. This transformation will result in a strain that is both resistant to the selectable marker but sensitive to 5% sucrose since the mutant cells convert the sucrose into levans, which are toxic fructose polymers (Dedonder, 1966). In the second transformation, the strain is transformed with a construct that contains only flanking sequences upstream and downstream of the inserted *sacB* and selectable marker genes. The transformation results in the creation of a markerless mutant (lacking the *sacB*-antibiotic resistance cassette) that can grow on 5% sucrose (Eaton-Rye, 2004, Liu and Curtiss, 2009, Vermaas, 2007, Ruffing, 2011). The *sacB* method for creating markerless mutants is illustrated in Figure 1.9. Such antibiotic-free strains are currently available for various model cyanobacterial strains.

Figure removed for copyright reasons

Figure 1.9: The *sacB* method for creating markless mutants.

(A) A DNA fragment carrying the selectable marker (SM)/*sacB* genes flanked on either end with regions homologous to the DNA sequences of the target site is integrated into the chromosome. (B) A DNA fragment carrying the desired deletion or insertion, again flanked by two regions homologous to the DNA sequences of the target sites, replaces the selectable marker */sacB* genes through homologous recombination. Modified from (Sun et al., 2008).

The potential uses of modern metabolic engineering tools and how they can play an important role in achieving large scale production of biofuels and other valuable products from microalgae were summarised by Chisti. (2007), and include: increase the photosynthetic efficiency of microorganisms to increase the biomass yield, speed up growth rates, increase the oil content in biomass, improve the tolerance of the biofuel producing organisms to variations in temperature to reduce the expense of cooling, eliminate the light saturation phenomena and reduce the susceptibility to photo-oxidation that leads to cell damage.

To sum up, the advancement of molecular genetic tools has allowed for the recent progress in cyanobacterial research and with time, the methods will be perfected further and new methods will be developed. However, it is important to combine genetic tools with other disciplines such as biochemistry in order to get more fruitful results. Furthermore, a better understanding of the system-level of microbial metabolism and the effects of product toxicity is also essential (Mukhopadhyay et al., 2008). Part of the work presented in this thesis covers the potential for tools developed aimed at rapid gene modification in *Synechocystis* where a potential method for random insertional mutagenesis is covered in chapter 3 and a simple and quick method for targeted gene deletions is discussed in chapter 4.

1.2 *Synechocystis* sp. strain PCC 6803

1.2.1 General overview

Synechocystis sp. strain PCC 6803 (hereafter, *Synechocystis*) is a freshwater, non-filamentous, non-nitrogen fixing, unicellular, coccoid cyanobacterium lacking gas vesicles or a sheath (Ikeuchi and Tabata, 2001). It was isolated from a freshwater lake in California in 1968 and is, as mentioned earlier, easily transformed by exogenous DNA by means of a well-established double homologous recombination system (Williams, 1988, Kufryk et al., 2002). It was the fourth living organism to have its genome fully sequenced (Kaneko et al., 1996) and the first photosynthetic autotrophic organism whose genome was completely sequenced and annotated (Dexter and Fu, 2009, Ikeuchi and Tabata, 2001). The genome consists of a single chromosome and seven plasmids of different sizes [four small and three larger (>5kb)] (Kaneko et al., 1996). The circular genome of this cyanobacterium was originally deduced to be 3.5 Mb (3.9 Mb including plasmids) with an average GC content of 47.7% (Koksharova, 2002, Kaneko et al., 1996) constituting 3317 assigned protein coding genes (3725 including plasmids-cyanobase) (Lu et al., 2011). Many researchers reported that *Synechocystis* species carry multiple genome copies per cell with most agreeing that it contains approximately ten copies (Eaton-Rye, 2004, Koksharova, 2002). A property shared with a number of other cyanobacteria (Vermaas, 1996). In terms of cell division, It was reported that the cells divide by binary fission in two or three successive planes (Ikeuchi and Tabata, 2001).

Although classified as a Gram negative bacterium, the DNA uptake mechanism during transformation is, overall, similar to that of Gram positive bacteria (Barten and Lill, 1995). Furthermore, *Synechocystis* cells, which are 1.5 to 2 μm in diameter (Ughy and Ajlani, 2004) (Figure 1.10), are much larger than average *E. coli* cells which are 0.5 μm in diameter.

Figure removed for copyright reasons

Figure 1.10: *Synechocystis* sp. strain PCC 6803.

(A) Wild type *Synechocystis* sp. strain PCC 6803 under the red channel showing laser excited chlorophyll (488 nm). Figure obtained from Joanna Sacharz (Queen Mary University of, London). (B) Wild type *Synechocystis* sp. Strain PCC 6803 under light microscope. Bar scale showing 5 μm in both images. Figure B was taken from the internet as detailed in Appendix 16.

Synechocystis is a mesophilic bacterium that grows optimally between 30°C and 34°C and at an optimal light intensity of 40-70 μE m^{-2} s^{-1} (Vermaas, 1996). It is a basophilic bacterium, hence grows well between pH 8.0 and pH 10.0 (Eaton-Rye, 2004). An interesting feature of *Synechocystis* is that it can utilise external fixed-carbon sources such as glucose or organic acids and grows very well in the presence of these compounds even if photosynthesis has been inhibited (Vermaas, 2007). That is probably why *Synechocystis* has a large variety of parallel physiological pathways. It has been reported that *Synechocystis* can grow fully photo-autotrophically, mixotrophically, and chemo-heterotrophically (Angermayr et al., 2009), a feature that enables the organism to survive and thrive in a wide range of physiological conditions (Vermaas, 2007). This property can be exploited to isolate and study viable mutants defective in photosynthesis (Astier et al., 1984). For this reason, this cyanobacterium has been widely used as a model for photosynthesis research (Eaton-Rye, 2004, Dexter and Fu, 2009). Table 1.3 summarises the different growth conditions in which the cyanobacterium can survive in and the average doubling times under each condition.

Table 1.3: Growth modes of *Synechocystis*. Adopted from (Vermaas, 1996).

Mode	requirements	Doubling time (h)
Photoautotrophic	CO_2,light,active PSII and PSI	13
Photoheterotrophic	glucose and active PSII or PSI	14 (with PSI only), 24 with PSII only
Photomixotrophic	glucose and active PSII or PSI	12
Heterotrophic	glucose; 5 min of light per day	28
Anaerobic	glucose or other fixed carbon source	approx. 48

Synechocystis are physiologically well characterised and relatively fast growing. Although Table 1.3 shows a minimum doubling time of 12 hours, some researchers have reported an average minimal doubling time of 6 to 12 hours (Heidorn et al., 2011, Angermayr et al., 2009). Furthermore, *Synechocystis* has no unusual nutritional demands, hence exhibits robust growth characteristics. In addition, the use of *Synechocystis* has the advantage of having a large body of methodological work associated with the genetics and genetic engineering strategies (Dexter and Fu, 2009). Its genome is genetically well characterised, and annotated, being publicly available at CyanoBase [1] and CyanoMutants [2] the two chief sites for information on cyanobacterial gene structure and gene function, respectively (Ikeuchi and Tabata, 2001). Finally, due to the development of genetic engineering tools for this model organism (section 1.1.4), there are a lot of data available on the regulation of gene expression and genetic designs related to promoter activities that result in controlled expression of heterologous metabolic pathways in this cyanobacterium (Angermayr et al., 2009). All these advantages make *Synechocystis* a well-established genetically-tractable platform for the production of many novel bio-products and an extremely useful and easy model to study cellular processes especially those of photosynthetic organisms.

1.2.2 Cell Structure & strain origins

Cell structure: since cyanobacteria are prokaryotic organisms, we expect their cellular organization to be similar to that of other bacteria. Indeed, cyanobacteria contain cell envelopes with combined features from both Gram positive and Gram negative bacteria. For instance, the overall structure resembles that of Gram negative bacteria but the peptidoglycan layer found in cyanobacteria is considerably thicker than that of most Gram negative bacteria (Hoiczyk and Hansel, 2000). In addition, the degree of crosslinking between the peptidoglycan chains within the cell wall layer of cyanobacteria (56–63%) is far higher than that in most Gram negative bacteria (20–33%). This may well be the reason why the

[1] http://www. kazusa.or.jp/cyano/

[2] http://www.kazuza.or.jp/cyano/mutants/

DNA uptake system in *Synechocystis* appears to be closer related to the one in Gram-positives (Barten and Lill, 1995). Furthermore, there are other components of cyanobacterial outer membranes that are not present in Gram negative bacteria such as carotenoids (Omata and Murata, 1984, Resch and Gibson, 1983), unusual fatty acids, such as β-hydroxypalmitic acid and porins (Hansel et al., 1998), to name a few.

The cell envelope of *Synechocystis* is generally composed of four layers: the external surface layers (such as S-layers and carbohydrate structures), the outer membrane, the polypeptidoglycan layer and the cytoplasmic membrane (Mullineaux, 1999, Liberton et al., 2006). These layers can clearly be seen in Figure 1.11. The S-layer commonly found in cyanobacteria is normally composed of a single species of (glyco)protein that covers the entire cell. It is believed to act as a protective coat or molecule sieve, or is possibly involved in cell adhesion and recognition (Sleytr et al., 1993).

Figure removed for copyright reasons

Figure 1.11: The envelope layers of the *Synechosystis* cell.

The S-layer (asterisk), outer membrane (white arrowhead), peptidoglycan layer (arrow), and cytoplasmic membrane (black arrowhead) are indicated. Scale bar, 50 nm. (Liu and Curtiss, 2009).

Strain origin: The Berkeley strain 6803, which was isolated from a freshwater lake in California in 1968 by R. Kunisawa (Stanier et al., 1971) is believed to be the origin of four other culture sub-strains ('PCC', 'ATCC', 'GT' (glucose-tolerant) and 'Kazusa'). Rippka R. (1992), explained that it was originally believed that these subcultures were the same. For this reason, they were grouped together under the name of *Synechocystis* sp. PCC strain number 6803. The four sub-strains were described by Ikeuchi and Tabata. (2001), where certain differences in their phenotypes were highlighted as are represented in Figure 1.12.

Figure removed for copyright reasons

Figure 1.12: Strain history of *Synechosystis* sp. PCC 6803 (Ikeuchi and Tabata, 2001).

Some of these differences are obvious phenotypic differences, while others are more subtle genetic variations (Williams, 1988). For example, the Berkeley and PCC strains are motile, while the others are apparently non-motile. Both PCC and ATCC strains are sensitive to glucose, whereas the glucose-tolerant (GT) strain was intentionally generated under defined conditions. Furthermore, the Kazusa strain is non-competent for transformation with exogenous DNA, while the other strains are naturally transformable (Grigorieva and Shestakov, 1982). It is important to note that only the Kazusa strain had its genome fully sequenced (Kaneko et al., 1996). In addition to the previous more detailed description of the sub cultures, the *Synechocystis* genus can generally be grouped into three groups based on the GC content; namely, the marine group, the low GC group and the high GC group, where *Synechocystis* PCC 6803 belongs to the latter group (Bergey and Holt, 1994). An interesting phenomenon that has been observed is that cultures has been found to change upon prolonged maintenance in the laboratory (Ikeuchi and Tabata, 2001), an important consideration to take on-board when working on strains to study mutations that result in a phenotypic difference.

1.2.3 Promoter systems

Suitable promoter elements (either native or foreign) are necessary for the effective and strong expression of foreign genes in any microorganism. Whether

the integration of the foreign gene takes place in the chromosome or on a replicating plasmid, the expression of a foreign gene requires a promoter that can be recognised in the host. Many native and foreign promoters have been evaluated in cyanobacteria mostly using *Synechocystis* 6803 and *Synechococcus* 7942 (Wang et al., 2012). In *Synechocystis*, several strong promoters have been identified (Ruffing, 2011). One example is the light regulated *psbA*II promoter that has been used to overexpress a number of native and foreign genes in *Synechocystis* (Lagarde et al., 2000, Lindberg et al., 2010, Mohamed et al., 1993). The fact that the expression of genes under this native promoter can be regulated is a big advantage. (Lindblad et al., 2012) showed that the foreign gene expression can be up regulated when cultures are grown under high light intensities and down regulated when cultured under low light intensities. Some genes, with products that are toxic to the cell, require a controlled expression and for that promoters such as the native nickel-inducible promoter *nrsB* can be utilised (Liu and Curtiss, 2009). The expression of the promoter is tightly regulated and functions in an on-off manner and also displays a dose-dependent response to added nickel, a very important feature for the regulated expression of certain genes. A list of promoters that have been used to drive expression in *Synechocystis* can be found in Table 1.4. It is clear from the table that the native promoters used are normally of genes that are essential to aspects of photosynthesis like CO_2 fixation (P_{rbcL}, P_{cmp}, P_{sbt}), photosystem I (PSI) (P_{psaA}, P_{psaD}), photosystem II (PSII) (P_{psbAI}, P_{psbA2}), and the photosynthesis antenna protein phycocyanin (P_{cpc}) (Wang et al., 2012). Besides native promoters, a number of foreign promoters have also been utilised in cyanobacteria, although the usage of such promoters has been limited. Interestingly to note is that although some strong *E. coli* promoters (*e.g.* P_{tac}/P_{trc}) have shown to express high levels of genes of interest in cyanobacteria, others (*e.g.* P_{lac}, P_{tet}, and λ P_R) show little or no detectable activity in *Synechocystis* (Wang et al., 2012). Furthermore, it has been reported that inducible promoters that are strictly regulated in species such as *E. coli* may not be controlled well in the cyanobacterial host (Elhai, 1994). All *Synechocystis* work conducted and presented in this thesis utilises the two native promoters *psbAII* and *nrsB* for their known strong and well regulated expression in this model organism.

Table 1.4: List of promoters used to drive expression of genes in *Synechocystis.*

Promoter	Source	Inducer/repressor and concentration	References
ArsB	*Synechocystis* sp. PCC 6803	Inducer AsO^{2-} (720 Mm)	(Blasi, Peca et al. 2012)
petE	*Synechocystis* sp. PCC 6803	Inducer Cu^{2+} (0.5 µM)	(Fernando, Verónica et al. 2012)
nrsB	*Synechocystis* sp. PCC 6803	Inducer Ni^{2+} (0.5 µM)	(Peca, Kos et al. 2007)
nrsB	*Synechocystis* sp. PCC 6803	Inducer Ni^{2+} (5 µM)	(Blasi, Peca et al. 2012)
nrsB	*Synechocystis* sp. PCC 6803	Inducer Ni^{2+} (6.4 µM)	(Peca, Kos et al. 2008)
nrsB	*Synechocystis* sp. PCC 6803	Inducer Ni^{2+} (15 µM) and Co^{2+}(3µM)	(Peca, Kos et al. 2007)
nrsB	*Synechocystis* sp. PCC 6803	Inducer Ni^{2+} (7µM and 20 µM)	(Liu and Curtiss 2009)
isiAB	*Synechocystis* sp. PCC 6803	Repressor Fe^{3+} (30µM)	(Kunert, Hagemann et al. 2000)
ZiaA	*Synechocystis* sp. PCC 6803	Inducer Cd^{2+} (2 µM)	(Blasi, Peca et al. 2012)
ZiaA	*Synechocystis* sp. PCC 6803	Inducer Zn^{2+} (5 µM)	(Peca, Kos et al. 2007)
ZiaA	*Synechocystis* sp. PCC 6803	Inducer Zn^{2+} (4 µM)	(Blasi, Peca et al. 2012)
ZiaA[a]	*Synechocystis* sp. PCC 6803	Inducer Zn^{2+} (3.5 µM)	(Berto, D'Adamo et al. 2011)
coat	*Synechocystis* sp. PCC 6803	Inducer Co^{2+} (6 µM)	(Fernando, Verónica et al. 2012)
coat	*Synechocystis* sp. PCC 6803	Inducer Co^{2+} (6.4 µM)	(Peca, Kos et al. 2008)
coat	*Synechocystis* sp. PCC 6803	Inducer Co^{2+} (3 µM)	(Peca, Kos et al. 2007)
coat	*Synechocystis* sp. PCC 6803	Inducer Co^{2+} (1 µM)	(Blasi, Peca et al. 2012)
coat	*Synechocystis* sp. PCC 6803	Inducer Zn^{2+} (3.2 µM)	(Peca, Kos et al. 2008)
coat	*Synechocystis* sp. PCC 6803	Inducer Zn^{2+} (5 µM)	(Peca, Kos et al. 2007)
coat	*Synechocystis* sp. PCC 6803	Inducer Zn^{2+} (4 µM)	(Blasi, Peca et al. 2012)
petJ	*Synechocystis* sp. PCC 6803	Repressor $CuSO_4$ (5µM)	(Tous, Vega-Palas et al. 2001)
psbA2	*Synechocystis* sp. PCC 6803	Inducer light (500µE/m^2/s)	(Lindberg, Park et al. 2010)
psbA2[a]	*Synechocystis* sp. PCC 6803	Inducer light (50µE/m^2/s)	(Berto, D'Adamo et al. 2011)
psbA2	*Synechocystis* sp. PCC 6803	Inducer light (100µE/m^2/s)	(Dexter and Fu 2009)
psaA	*Synechocystis* sp. PCC 6803	Inducer light (20 µE/m^2/s)	(Muramatsu and Hihara 2006)

psaD	*Synechocystis* sp. PCC 6803	Inducer light (20µE/m^2/s)	(Muramatsu and Hihara 2007)
rnpB	*Synechocystis* sp. PCC 6803	-	(Huang, Camsund et al. 2010)
cmpA	*Synechocystis* sp. PCC 6803	inducer CO_2 limitation	(Liu, Fallon et al. 2011)
sbtA	*Synechocystis* sp. PCC 6803	inducer CO_2 limitation	(Liu, Fallon et al. 2011)
CI-PL	*Synechocystis* sp. PCC 6803	Temp inducible	(Dexter and Fu 2009)
A1lacO-1	*E. coli*	Inducer IPTG (1mM)	(Fernando, Verónica et al. 2012)
trc20	*E. coli*	Inducer IPTG (2mM)	(Huang, Camsund et al. 2010)
trc10	*E. coli*	Inducer IPTG (2mM)	(Huang, Camsund et al. 2010)
Trc	*E. coli*	Inducer IPTG (1mM)	(Fernando, Verónica et al. 2012)
tet	*E. coli*	-	(Huang, Camsund et al. 2010)
nirA	*Synechococcus* PCC 7942	Inducer/Repressor NO_3-/NH_4+ (17.6mM/17.6mM)	(Qi, Hao et al. 2005)

1.2.4 Plasmid vectors used for *Synechocystis*

Plasmid vectors that are used in transformation of cyanobacteria can be classified into two main groups according to their mode of replication in the host (Koksharova, 2002): the first group is non-replicating plasmids or integrative plasmids and these cannot replicate independently and would eventually be lost through successive cell divisions (Wang et al., 2012). The other group is commonly referred to as replicative plasmids, shuttle vectors or biphasic plasmids and are capable of replication in both cyanobacteria and *E. coli* (Deng and Coleman, 1999, Koksharova, 2002, Wang et al., 2012) as they most often contain replication sequences from both *E. coli* and the cyanobacterial host (Ruffing, 2011). Both integrative and replicative plasmids have been developed for cyanobacteria and have their experimental uses. For example, the ideal plasmid vector for transporting DNA that must either transpose (*e.g.* a transposon) or integrate (*e.g.* by homologous recombination) in order to be stably maintained are the integrative vectors (Eaton-Rye, 2004, Golden et al., 1987). Hence, these vectors are ideal for targeted mutagenesis in which the gene in the chromosome is replaced by the mutated gene in the plasmid (Golden et al.,

1987) or in which the cyanobacterial host needs to select for the recombination of the plasmid into the chromosome (Williams and Szalay, 1983). Interestingly, it has been observed that none of the common cloning vectors from *E. coli* seem to be able to replicate in cyanobacteria (Porter, 1986), presumably because the colE1 origin of replication carried on vectors such as pUC19, pBluescript and pGEM-T are not recognised by the DNA replication machinery of cyanobacteria (Eaton-Rye, 2004).

Replicative plasmids can be classified into two types: those that contain replicons of broad-host range plasmids (Ng et al., 2000, Huang et al., 2010) and those derived from endogenous cryptic plasmids (Deng and Coleman, 1999, Wolk et al., 1984). Replicative plasmids carrying a variety of selectable markers, have been constructed for cyanobacteria (Golden et al., 1987). Examples include the pFCLV 7 (Chauvat et al., 1986) and plasmids derived from RSF1010 (for example, pSL1211, pPMQAK1 and pFC1) (Berla et al., 2013, Ng et al., 2000, Wang et al., 2012). Replicative plasmids that can replicate in multiple hosts (either broad or narrow host range) have the potential to express an exogenous gene in numerous hosts (Koksharova, 2002). Nevertheless, most shuttle vectors for cyanobacteria are still not well characterised for their copy number (Ruffing, 2011), and the lack of shuttle vectors with a varied copy number puts limitations on the controlled expression of heterologously expressed genes by selecting their copy number (Dunlop et al., 2011). For example, the copy number from plasmids derived from RSF1010 are 10 per chromosome in *E. coli* cells and approximately 10-30 per cell in *Synechocystis* (~1-3 per chromosome) (Huang et al., 2010, Ng et al., 2000). However, the copy number of other broad host-range plasmids have not been quantified to date (Berla et al., 2013), which adds a limitation to their use. A recent review is available on shuttle vectors used in various cyanobacterial hosts (Wang et al., 2012). An illustration of the integrative and replicative plasmids can be found in Figure 1.13.

It is interesting to note that transformation of *Synechocystis* with plasmids obtained from *Synechocystis* cultures result in 10-20 times better transformation rates than transformations done using bacterial plasmids, perhaps due to a compatible DNA methylation pattern. Nevertheless, plasmids extracted from *E. coli* are routinely used as they are much more convenient to prepare (Labarre et al., 1989).

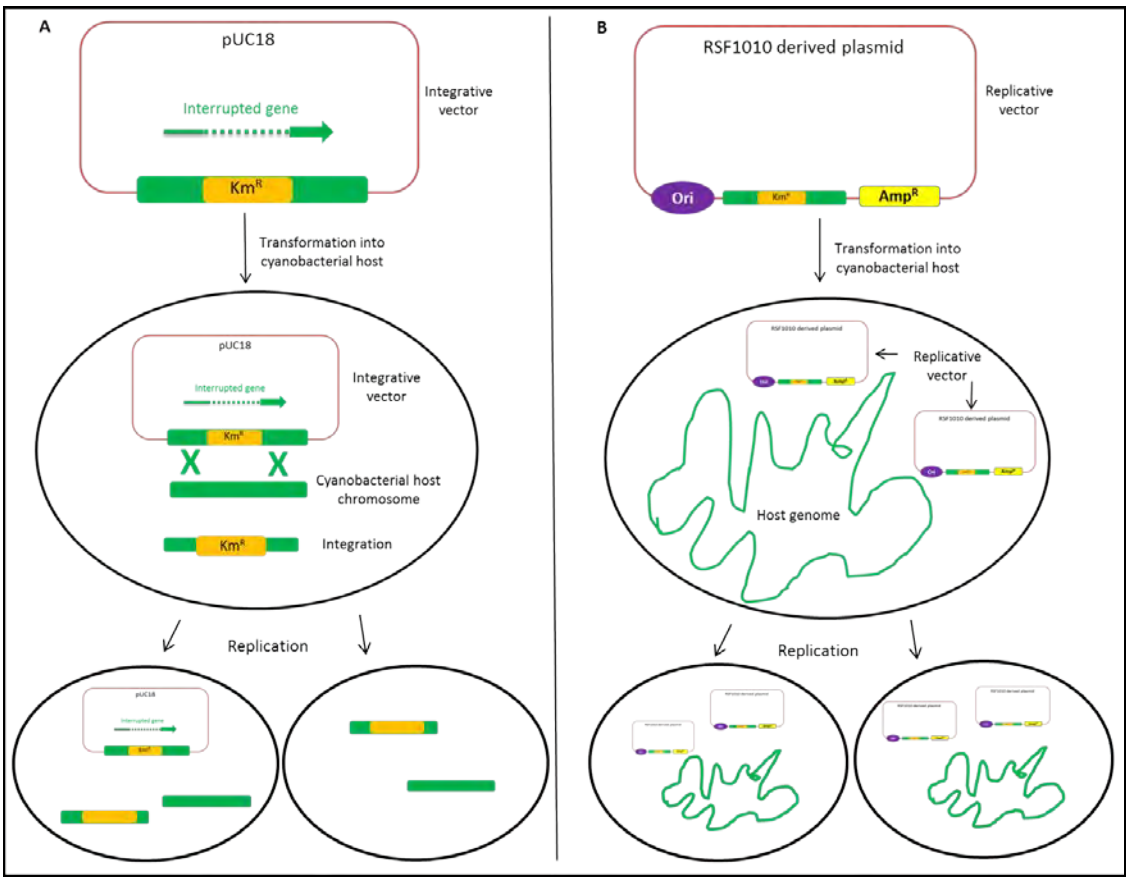


Figure 1.13: An illustration of the plasmids used for cyanobacterial transformation

(A) Example of an integrative plasmid, the plasmid is eventually lost upon successive replication (B) An example of a replicative plasmid showing maintenance of the plasmids upon replication. **Key**: Km^R: kanamycin resistance gene, Amp^R: ampicilone resistance, Ori: origion of replication.

1.2.5 Selectable markers used in *Synechocystis*

It is essential to be able to select for rare transformant events amongst a large population of treated cyanobacterial cells (Vermaas, 2007, Porter, 1986), hence systematic and efficient methods whereby successful transformants can be identified and isolated are well established. Usually, host cells that contain the transferred DNA are identified by transferring a gene whose product has the potential to inactivate an antimetabolite, such as an antibiotic (Koksharova, 2002). When cultured in the presence of the antimetabolite, only those cells to which the gene has been transferred should grow. Since cyanobacteria are Gram-negative bacteria, antibiotics and antibiotic-resistance genes that are known to be effective for that group of bacteria have been exploited extensively (Vermaas, 2007).

Antibiotics used as selectable markers in cyanobacteria, in general, include neomycin and kanamycin and the corresponding neomycin phosphotransferase

(npt)-encoding genes from transposons Tn5 and Tn903, and streptomycin and spectinomycin and the corresponding aminoglycoside adenyltransferase (aadA)-encoding genes from Tn7 (Wolk et al., 1993) and the omega interposon (Prentki et al., 1991). Bleomycin, although expensive, is sufficiently active to be cost effective with *Anabaena* 7120. Chloramphenicol and chloramphenicol acetyl transferase (cat) have proven useful for unicellular cyanobacteria (Chauvat et al., 1986) and *Anabaena* sp. strain 90 (Rouhiainen et al., 2000). Antibiotics such as tetracycline and rifampicin, which are light-sensitive, are of much reduced usefulness for cyanobacteria (Koksharova, 2002, Ruffing, 2011). Other antibiotics with the same property include: erythromycin (Vermaas, 1996), zeocin (Howitt et al., 1999) and gentamicin, (He et al., 1999).

Synechosystis shows preference to some antibiotics. For example, it has been reported that the transformants on the plates containing kanamycin grew 4-6 days earlier than those on the plates containing chloromycetin (Zang et al., 2007). Furthermore, the bacterial promoter and termination sequences of kanamycin is recognised by *Synechosystis* and hence can be used without further modifications (Vermaas, 2007). A summary of the selectable markers (antibiotics resistance markers) suitable for use in *Synechocystis* can be found in Table 1.5.

By using a combination of antibiotic resistant cassettes identified in Table 1.5 multiple genes can be inactivated in a single *Synechosystis* strain (Jansson et al., 1987, Chu et al., 1994, Morgan et al., 1998). One thing to keep in mind when using multiple antibiotic resistant cassettes is that the cassette normally used for spectinomycin resistance also results in streptomycin-resistance. Therefore, it is necessary to introduce the streptomycin-resistance cassette before the spectinomycin resistance cassette (Vermaas, 2007).

Table 1.5: Selectable markers (antibiotics resistance markers) suitable for use in *Synechocystis*.

Selectable marker	Source	concentration (µg/ml)	References
chloramphenicol	pBR325	5-150	(Vermaas, 1998, Eaton-Rye, 2004)
erythromycin	pRL425	5 -300	(Vermaas 1998)
kanamycin	pUC4K (Tn 903)	5-500	(Vermaas, 1998, Eaton-Rye, 2004)
spectinomycin	pHP45Ω	3-250	(Golden et al., 1987, Vermaas, 1998)
neomycin	Tn5	-	(Koksharova, 2002)
streptomycin	Tn7	-	(Wolk et al., 1993)
spectinomycin	pHP45Ω (Tn7)	25	(Eaton-Rye 2004)(Wolk et al. 1993)
zeocin	-	25	(Howitt, Udall et al. 1999)
gentamicin	-	-	(He, Schlich et al. 1999)
SacB	*Bacillus subtilis*	5mM sucrose	(Vermaas 2007)

As mentioned in section 1.1.4, the creation of markerless deletions of genetically modified organisms is critical for industrial applications and can be achieved in *Synechocystis* using the *sacB* gene since its expression has shown to be lethal when sucrose is present in the medium (Ermakova-Gerdes and Vermaas, 1999). A recently discovered system for the creation of markerless mutants in *Synechocystis* involves the use of the *E. coli maz*F gene, a protein synthase inhibitor expressed under a nickel inducible promoter, which operates in a similar manner as the toxic *sacB* gene described in Figure 1.9 (Cheah et al., 2013). Since both systems are based on negative selection in the second transformation, care must be taken to ensure that the loss of sucrose resistance is due to the counter-selectable marker and not due to mutation. However, either method allows for the re-use of a single selectable marker for making multiple successive modifications to the host chromosome (Berla et al., 2013).

1.2 Biofuels

The production of biomass-derived fuels is gaining increasing interest worldwide. Climate change, rising fuel prices, concerns about greenhouse gas emissions and political instability in the producing countries are some of the recent concerns that sparked interest into alternatives for fossil fuels (Al-Haj, 2012). Currently, fuels make up approximately 70% of the global final energy market (Parmar et

al., 2011) while the remaining 30% is accounted from global electricity (Hankamer et al., 2007). The US enacted the Independence and Security Act of 2007 where it set a goal of producing 36 billion gallons of renewable fuel by 2022 with 16 billion to be obtained from cellulosic ethanol (EISA, 2007). According to the Renewable Fuels Association, 80 billion litres of biofuels were produced worldwide in 2009 most of which were crop-based fuels (bioethanol from corn and biodiesel from soybeans) (Zhou and Li, 2010).There are on-going debates over the most suitable host for biofuel production, the best type of biofuel, and the best biological platform to use for biofuel production. These topics along with challenges facing biofuel production will be discussed in the following sections.

1.3.1 Choosing a biofuel-producing host

There are basically three alternative routes to convert renewable resources into energy-rich, fuel-like molecules or fuel precursors (Rude and Schirmer, 2009): first, fermentative or non-fermentative production by heterotrophic microorganisms, such as bacteria, yeast or fungi; second, direct production by photosynthetic organisms, such as plants and algae; and third, chemical conversion of biomass to fuels (not covered here). It is therefore essential to identify the pros and cons of producing biofuels from heterotrophic and phototrophic microorganisms in order to decide which host to use for a novel microbial biofuel production.

It is evident that the two heterotrophic model organisms *Saccharomyces cerevisiae* and *Escherichia coli* are being used to develop most of the new microbial fuels (Lee et al., 2008). Needless to mention, these model organisms have been used in industrial processes for centuries for the production of fuels, chemicals, and pharmaceuticals (Alper and Stephanopoulos, 2009). Nevertheless, the main drawback of using these heterotrophic microbes is the cost of their organic feedstock (carbohydrates), which remains the major cost driver for fuels production (Rude and Schirmer, 2009, Ducat et al., 2011b). Another drawback is the recovery cost associated with the fuel extraction from the fermentation broth and the cost for any additional chemical modifications necessary to convert a precursor fuel into an engine-ready fuel. It would be ideal to generate microorganisms that can produce biofuels similar to current petroleum-based transportation fuels as it would allow the use of existing engines

and infrastructure and would save a vast amount of capital necessary for replacing the current infrastructure to accommodate biofuels that have properties significantly different from petroleum-based fuels (Lee et al., 2008).

Plants, on the first view, look more attractive as they produce fuels directly from CO_2 using photosynthetic metabolism; hence the process is carbon-neutral and does not rely on expensive feedstocks. However, this approach also faces several major hurdles, most importantly the use of a food source for fuel, scalability, the need for freshwater resources, and land use issues (Al-Haj, 2012). Unicellular microalgae and cyanobacteria on the other hand, seem to have an advantage for biofuels relative to terrestrial plants, some of which are summarised by Dismukes et al. (2008), in the following points:-

Microalgae require a smaller land footprint and less irrigation for cultivation, many species of algae thrive in salt water and are able to grow all year round in diverse conditions. Furthermore, they do not accumulate recalcitrant lignocellulosic biomass and more importantly, genetic manipulation techniques have been developed for some species, and are increasingly being applied to optimise biofuel production in several algal systems. Furthermore, photosynthetic microorganisms, such as cyanobacteria and unicellular algae, are more efficient than crop plants in converting solar energy to biofuels (Melis, 2009, Beer et al., 2009). For example, microalgae has shown to convert up to 3% of solar energy into chemicals while higher plants a mere 0.5% (Li et al., 2008). Another advantage of photosynthetic organisms is that they can grow to high densities within a fully enclosed photo-bioreactor (Chisti, 2007, Beer et al., 2009, Angermayr et al., 2009) and can tolerate higher CO_2 content in gas streams, an attractive advantage for the conversion of CO_2 from flue gases into organic substances (Zhou and Li, 2010). Cyanobacteria offer yet another advantage over algae and that is faster growth rates and easier genetic manipulation (Berla et al., 2013).

Oleaginous microbes (those that accumulate >20% of their cellular dry weight in lipids) whether phototrophic or heterotrophic, can be used as another option to achieve high yields of microbial fuels (Rude and Schirmer, 2009). For example, yeast, fungi and algae are being utilised as they accumulate high amounts of lipids in the form of triacylglycerol (TAG) (Hu et al., 2008, Li et al., 2008). In some reports, the amount of TAGs accumulated as lipid storage during microbial stress

can reach 20-50% of the cell dry weight. However, naturally occurring cyanobacteria have not been shown to accumulate significant amounts of TAGs (Hu et al., 2008). Nevertheless, this problem can probably be mitigated by utilising advances in metabolic engineering and synthetic biology to convert model cyanobacterial strains such as *Synechocystis* or even heterotrophic organisms such as *Saccharomyces cerevisiae* and *Escherichia coli* into oleaginous organisms (Rude and Schirmer, 2009). At the moment, it is not clear whether using a single organism or a consortium is best for biofuel production. The question whether a single organism can be engineered that out-competes a collection of specialised organisms has yet to be answered (Dismukes et al., 2008).

The development of a “chassis” organism for synthetic biology is a difficult task (Lee et al., 2008). Ideally, the microorganism of choice for biofuel production should be able to utilise high amounts of substrate, accumulate high amounts of lipid, able to process it quickly, should have high metabolic fluxes, produce a single product and have good tolerance to inhibitors and product (Alper and Stephanopoulos, 2009). The organism should also be able to grow relatively fast and be amendable to genetic manipulation. Using cyanobacteria as the host of choice for biofuel production is an attractive option for the many advantages they offer. Nevertheless, direct biofuel production from cyanobacteria still faces some bottlenecks and until today no ideal chassis organism that ticks all the boxes has been found or created. Whether such an organism will be engineered using an isolated strain or an established model organism as the starting point is unclear. The choice between engineering natural function and importing biosynthetic capacity depends on the current progress in metabolic engineering and synthetic biology (Alper and Stephanopoulos, 2009).

In summary, after analysing the different parameters involved in selecting a host organism for biofuel production it seems clear that employing engineered photosynthetic microbes, specifically cyanobacteria strains, is the most efficient strategy to produce novel biofuels. Recent advances in molecular, systems, and synthetic biology now allow for the rapid engineering of microbial biosynthetic pathways to produce a variety of advanced biofuel candidates such as alcohols, esters, alkanes, and alkenes from the isoprenoid and fatty acid pathways (Peralta-Yahya and Keasling, 2010). The ideal microorganism that consolidates

all desirable biofuel processing characteristics does not exist so far (Alper and Stephanopoulos, 2009). Some of the desired individual characteristics mentioned above can be found in certain species but never combined in one “super strain”. Finally, one must be aware that in the pursuit of renewable biofuels, we need not only look at the best “organism” for biofuel production but also the “process” and “product” (Lindberg et al., 2010).

1.3.2 Types of biofuels

In order to replace current liquid transport fuels with biofuels, alternatives to gasoline, diesel and jet fuel need to be found (Peralta-Yahya and Keasling, 2010). There are certain physical and chemical characteristics that a biofuel needs to meet in order to replace such fuels. Therefore, it is critical to assess the produced biofuels and determine their suitability as fuel molecules (Fortman et al., 2008). A recent publication investigated the potential of producing isoprenoid-derived fuels in cyanobacteria alongside terpene combustion characterisation and emissions produced in both gasoline and diesel engines (Hellier et al., 2013). Such studies are critical for the development of biofuels that are sustainable in both production and utilisation. For example, replacement fuels need to be stable in the engine and show predictable combustion at high pressures as well safe to store and transport, and of course contain high energy density (Antoni et al., 2007). They also need to show similar chemical composition to fossil fuels in order to be compatible in the existing engines and transportations systems (Hellier et al., 2013). Fatty acid esters, long chain fatty alcohols and fatty alkanes are a few examples of such fuels (Lu, 2010). More examples of fuels that meet some/all of these requirements and the fossil fuel types they can replace either as full replacement fuels or fuel blends are found in Table 1.6. Researchers engaged in finding fossil fuel replacements such as those listed in Table 1.6 have recently produced fermentative alcohols (Atsumi et al., 2008a, Hanai et al., 2007, Inui et al., 2008, Nielsen et al., 2009, Steen et al., 2008, Jojima et al., 2008), non-fermentative higher alcohols (Atsumi et al., 2008b), isoprenoid-derived fuels (Carter et al., 2003, Withers et al., 2007) and lipid fuels (Steen et al., 2010, Lu et al., 2008, Liu et al., 2010, Kalscheuer et al., 2006). Representative structures of the different fuel types can be found in Figure 1.14. Alternative biofuels can be of a gaseous or liquid nature and each exhibits its pros and cons. Fuels that are

liquid at atmospheric pressure and ambient temperature (*e.g.* alcohols) are easier to store, distribute and use while gases from microbial origin (*e.g.* hydrogen, methane, and isoprene), although easer to capture from the gas space of culture vessels, need to first be liquefied by cooling and compression for transport (Lu et al., 2011). Furthermore, they are more difficult to store and transport and would probably need a new distribution infrastructure. Liquid biofuels however, need to remain in the liquid form and be pumpable at all temperatures encountered and remain stable during storage (Antoni et al., 2007).

Table 1.6: Biofuel alternatives to the conventional transport fuels.

conventional fuel	advanced biofuel alternative	carbon length
Gasoline	short -chain alcohols and alkanes	C_4-C_{12}
Diesel	FAMEs, fatty alcohols, alkanes and linear or cyclic isoprenoids	C_9-C_{23}
Jet fuel	fatty acid and isoprenoid based fuels	C_8-C_{16}

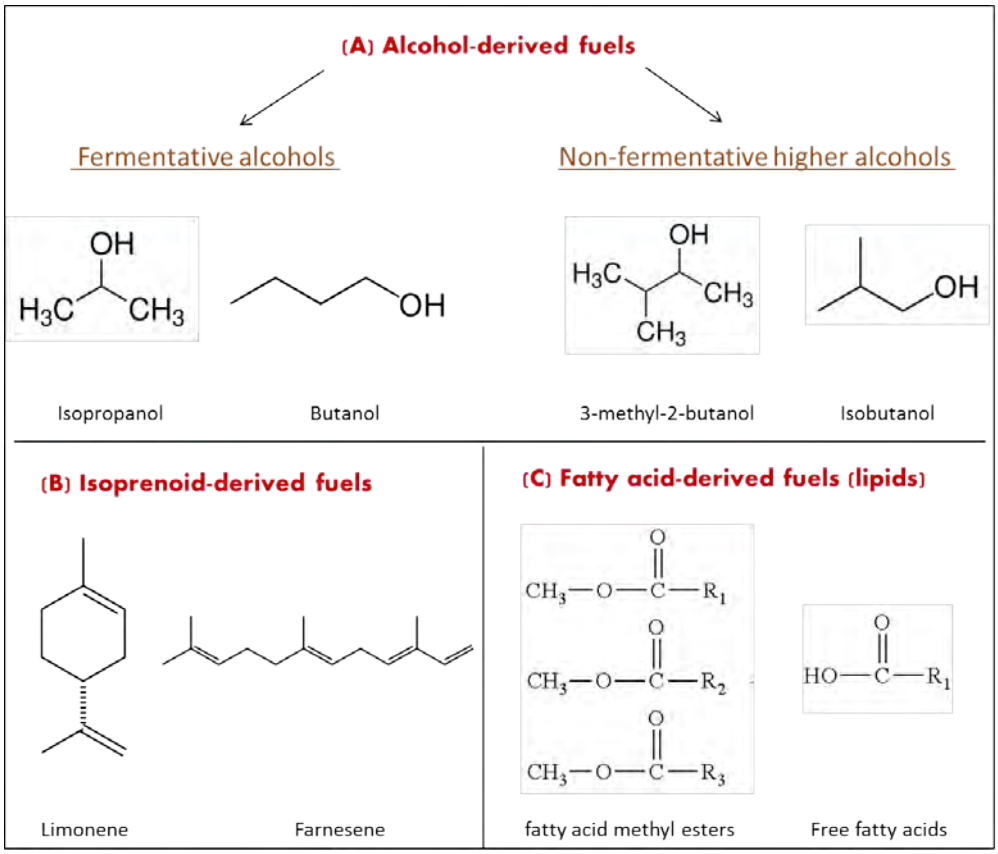


Figure 1.14: Representative structures of the different fuel types

(A) Examples of alcohol-derived fuels including short-chain alcohols and higher alcohols, (B) Examples of Isoprenoid-derived fuels and (C) Examples of lipid-based fuels.

Out of the variety of biofuel options available, only biodiesel and bioethanol are produced in industrial scale today, biomethane is produced in a large scale but still not utilised for transport, and biobutanol is gaining increasing interest and seems that it will soon be used as a supplement for gasoline, diesel and kerosene while both biohydrogen and biomethanol need further development (Antoni et al., 2007). Despite the recent hyperbole on biofuels, the actual concept of biofuel is surprisingly old and was used throughout man's long history. Rudolf

Diesel, who is the inventor of the diesel engine, used vegetable oil as a fuel source for his engine in the 1900's [3]. Before that, in the 1860's, Nikolaus August Otto used ethanol in his prototype of the spark ignition engine (gasoline replacement). His work was sponsored by a sugar factory interested in mass production of ethanol (Antoni et al., 2007). The use of vegetable oil (*e.g.* peanut oil) along with corn based fuel were used for a long time until petroleum entered the picture and proved to be a superior fuel. This was determined from factors such as the fuel supply, fuel price and efficiency. Nevertheless, both vegetable oils and ethanol continued to be used as diesel and gasoline replacement fuels during the 1930s and 1940s although at a much smaller scale (Antoni et al., 2007). The concept of using algae for biofuels was introduced in 1955 by Meier (1955), who used the carbohydrate fraction of the cell for methane production. The idea was further improved in 1960 when growing algae in large raceway ponds and digesting the algal biomass for methane production was introduced (Oswald and Golueke, 1960). In the 1970's, although the concept of ethanol revived due to the "crude oil crisis" (Antoni et al., 2007), the potential use of algae as biofuels gained more attention as concerns for raising conventional fuel prices increased (Hu et al., 2008). Since then microbial systems specifically have been extensively explored and successfully used for the biosynthesis of some biofuels, most notably ethanol. However, with increased knowledge of the properties required for ideal biofuel molecules, scientists agree that ethanol might not be the best option as it is corrosive and cannot be used in the current distribution infrastructure (Connor and Atsumi, 2010).

Furthermore, higher-chain alcohols such as isobutanol offer several advantages compared with ethanol, such as higher energy density, lower water solubility (Atsumi et al., 2008b) and less corrosion (Ducat et al., 2011a). Algal-derived biofuels currently being investigated include alcohols from starches, biodiesel from lipids and H_2 for fuel cells (Beer et al., 2009). However, thus far, crop-derived biodiesel and bioethanol (Ragauskas et al., 2006, Lee et al., 2008) and algal biodiesels (Chisti, 2007) are the main biofuels being produced at an industrial scale with algal biodiesels considered the most efficient (Chisti, 2007).

---

[3] http:// www.biodiesel.com/index.php/biodiesel/history_of_biodiesel_fuel.

Nevertheless, the many intermediate steps involved in the production of both types of fuels contribute significantly to the high cost of production (Atsumi et al., 2009) and it is therefore essential to find means to achieve a more cost effective industrial scale production.

Biofuels can be classified into first, second, third and fourth generation according to the feedstock. Biofuels that are made from crop plants such as corn and sugar cane, discussed above, are called first generation biofuels. While second generation refers to the utilisation of all forms of (ligno)cellulosic biomass for biofuels production and includes non-crop plants such as willow and *Miscanthus*. Both third and fourth generation biofuels apply to the utilisation of microalgae (algae and cyanobacteria) for the " algae-to-fuel" technology; where the former uses the actual algae biomass for biofuel production and the latter employs metabolic engineering for the modification of microalgae in order to convert them into microbial fuel factories (Lu et al., 2011). The four generations of biofuels will be discussed in more detail in section 1.3.3.

1.3.3 The four generations of biofuels

Many platforms have been explored as a source for biofuel production ranging from plants, to waste organic material, to microorganisms and microalgae (Al-Haj, 2012). As introduced in the previous section, the four generations of biofuel differ mainly in the biomass resource platform.

First generation biofuels: utilises plant-derived biomass such as corn grain, sugarcane, soybean and rapeseed for biofuel production (Dismukes et al., 2008). Examples of biofuels produced from plant-derived biomass include bioethanol, biomethane and biodiesel. Besides the obvious food for fuel issue associated with first generation biofuels [e.g. the US currently uses 25% of its total corn production for biofuels (Zhou and Li, 2010)], first generation biofuels suffer many other drawbacks.

Traditionally bioethanol was produced via yeast fermentation using relatively expensive feedstocks, such as corn starch and cane sugar (Deng and Coleman, 1999). Bioethanol production from sugarcane was pioneered in Brazil (Goldemberg, 2007) and currently bioethanol from starch is the main type of first generation biofuels (Lu, 2010) and the major biofuel being used in the transportation sector (Peralta-Yahya and Keasling, 2010). Bioethanol has many

disadvantages such as incompatibility with the distribution and storage infrastructure, corrosiveness, low energy content (Fortman et al., 2008), is based on the limited agricultural/food resources and needs large agriculture land (Sheehan, 2009, Hellingwerf and Teixeira de Mattos, 2009, Dexter and Fu, 2009), furthermore due to its high volatility, it is difficult to transport and finally it is toxic to the producing host (Lu, 2010). The production of biomethane is not a new concept and has been achieved decades ago (Hellingwerf and Teixeira de Mattos, 2009). Although the large-scale production of biomethane as a biofuel has been largely overlooked in the past 20 years (Åhman, 2010), it remains of interest today (González-Fernández et al., 2012, Murphy and Power, 2009, Tilche and Galatola, 2008, Åhman, 2010). However, being a gaseous fuel, it has disadvantages. Moreover, the conventional production of biomethane from woody biomass via gasification needs to be further developed (Åhman, 2010). Biodiesel on the other hand, is produced mainly from soybeans (Chisti, 2007) while other sources for commercial biodiesel include canola oil, animal fat, palm oil, corn oil, waste cooking oil (Felizardo et al., 2006, Kulkarni and Dalai, 2006) and jatropha oil (Barnwal and Sharma, 2005). Biodiesel offers several advantages over bioethanol and biomethane, most notability nontoxicity and ease of storage and transport (Schenk et al., 2008). Furthermore, biodiesel production is cheaper than bioethanol production in some cases (*e.g.* the energy yield from soy oil per hectare is far lower than the ethanol yield from corn) (Pimentel and Patzek, 2005). Nonetheless, scientists researching into biofuels are turning away from ethanol derived from sugarcane and starch and away from biodiesel derived from animal fats and vegetable oil, and are finding alternatives to these crop based biofuels (Sheehan, 2009).

Second generation biofuels: These include plant-derived lignocellulose from agriculture and agro-industrial residues. Examples of which are wood chip, corn stover, sugarcane, rice and wheat straw (Machado and Atsumi, 2012) and in some cases, plant seeds as many seeds accumulate high amounts of tri-glycerides, which upon transestrification produce biodiesel (long-chain fatty acid methyl esters) (Hellingwerf and Teixeira de Mattos, 2009). Second generation biofuels are therefore not tied to food crops or in limited supply. However, the use of lignocellulosic biomass can be expensive since this biomass is tough due to the presence of cellulose, hemicellulose and lignin that needs to be enzymatically

processed (Lu et al., 2011). The discovery of enzymes that can degrade plant polysaccharides and lignins into simple sugars available for microorganisms to ferment into fuel would open the door for a more efficient conversion of lignocelluloses to biofuels. There is much research focused on solving this issue, however the solution is a long term goal (Sanderson, 2011).

Third generation biofuels: Until recently, biodiesel was only produced from plant and animal oils and not from microalgae (Chisti, 2007). However, this is changing rapidly as companies are attempting to commercialise microalgal biodiesel (Solazyme, Synthetic Genomics and Joule to name a few). Third generation biofuels includes the production of natural oils (lipids), from algae, for the conversion into some form of renewable biodiesel (Sheehan, 2009). A process that requires growth, harvesting and extraction (Chisti, 2007). For some species the natural content of lipids is ~30% by weight and for selected species this content can go up to 80% (*e.g. Schizochytrium* sp and *Botryococus braunii*) (Hellingwerf and Teixeira de Mattos, 2009). The lipid content in microalgae can be increased by subjecting them to stressful environmental conditions. This has been demonstrated by several researchers. For example Liu et al. (2008), showed that using high initial concentrations of iron in the culture media elevated lipid accumulation in the microalga *Chlorella vulgaris,* while Illman et al. (2000), showed that nitrogen depravation led to the accumulation of total lipids in five different *Chlorella* strains. Third generation biofuels represented in algal-derived fuels offer many competitive advantages over first and second generation biofuels. For example, algae can grow extremely rapidly [doubling times of <24 hours and in some cases as fast as 3.5 hours per day (Chisti, 2007), they are not based on food and do not require arable land (Parmar et al., 2011) and can utilise a wide variety of water sources (Hu et al., 2008). Some microalgae accumulate large quantities of neutral lipids and can thrive in adverse ecological conditions (Hu et al., 2008). Furthermore, algae are not only utilised for their oils but also the remaining biomass can be converted to biogas through anaerobic fermentation (Wijffels and Barbosa, 2010). It is for all these reasons that algal biodiesel has been a hot topic in biofuel production recently. Nevertheless, producing microalgae biomass is generally more expensive than growing crops since they need light, CO_2, water, inorganic salts such as nitrogen, phosphorus, iron and some cases silicon, and need to be maintained at temperatures around 20-30^{o}C

(Chisti, 2007). Furthermore, the production of microbial fuels requires high energy inputs due to the many steps involved in microalgal harvesting (including flocculation, sedimentation, flotation, centrifugation or filtration). There are also other costs associated with the actual lipid extraction where harsh conditions (*e.g.* mechanical, chemical and physical stress) need to be used to break the thick cell walls that algae normally possess (Wijffels and Barbosa, 2010, Sheehan, 2009). Some of these issues are currently being addressed by developing what is called "Fourth generation biofuels".

Fourth generation biofuels: In the development of fourth generation biofuels, also referred to as advanced biofuels, advances in the field of science and engineering need to be employed to ensure the produced fuels are sustainable in both production and utilisation (Hellier et al., 2013). Fourth generation biofuels involves the employment of synthetic biology and genetic engineering to modify photosynthetic microorganisms in order to turn them into factories for biofuel production (Dexter and Fu, 2009, Lu et al., 2011). The main difference between 4^{th} and 3^{rd} generation biofuels is that the former applies the "cell factory" concept for the production of gaseous and liquid biofuels by utilising advances in synthetic biology and genetic engineering approaches to both produce and excrete the biofuel out of the cell, thereby reducing the cost of downstream processing (Lu et al., 2011). As mentioned in section (1.1.2), engineering cyanobacterial strains by means of either single gene modification (deletions/insertions) or pathway engineering (native or heterologous) could lead to production of novel fuel molecules (Table 1.1). Fourth generation biofuels that are currently being produced include long chain alcohols, isoprenoid derived fuels and fatty-acid derived fuels (Peralta-Yahya and Keasling, 2010). Examples of advanced biofuels that have recently been produced in engineered cyanobacteria are listed in Table 1.1. Although several potential advanced biofuels have been produced in microbes, metabolic engineering tools need to be employed to increase the titre for the cost effective industrial production of these fuels (Peralta-Yahya and Keasling, 2010). Furthermore, there are many factors that need to be considered when designing novel fuels and these include: the energy content, combustion quality, volatility, freezing and boiling points, odour and toxicity (Lee et al., 2008). The aim is to produce advanced biofuels with high energy content that are compatible with the exciting storage and distribution infrastructures, and that are

also economically feasible (Peralta-Yahya and Keasling, 2010) while keeping in mind that metabolically engineered strains are forced to produce a compound that my lead to its autoinhibition (Angermayr et al., 2009).

A main advantage of 4th generation biofuels is the reduced number of steps needed to go from sunlight capture to formation of fungible biofuels. It is clear in Figure 1.15 that the steps needed are much less compared to the previous methods thereby reducing the large capital costs associated with production. However, the development of 4th generation biofuels is still in its infancy and the technology lags far behind that established for the other methods (Lu et al., 2011).

At this point no advanced biofuel production technology has come out on top. However, with the wide range of molecules in petroleum-based fuels and the large amount of renewable fuels urgently needed to power the large transportation sectors, there is enough room to accommodate all technologies. The development of advanced fuels that will work with the existing transportation infrastructure is certainly a step in the right direction toward usability and economic viability (Peralta-Yahya and Keasling, 2010).

Figure removed for copyright reasons

Figure 1.15: Comparison fo the typical bioprocess steps needed for 4 generations of biofuels production

It is clear how 4th generation processes shorten the steps towards biofuel production significantly. Microalgae refer to both algae and cyanobacteria. Figure modified from (Lu et al., 2011).

1.3.4 Challenges facing biofuel production

Besides challenges associated with the selection of the best type of biofuel or the most suitable host for biofuel production discussed in the previous sections, there are a lot of other challenges that need to be overcome for successful, commercial level, biofuel production. For example, biofuel production yields need to be enhanced in a cost effective manner in order to become competitive to fossil fuels. Issues related to fuel toxicity need to be addressed (Zhou and Li, 2010) and the availability of metabolic cofactors required in engineered microbial pathways need to be enhanced. Furthermore, product scale–up and biorefineries need to be set up (Wijffels and Barbosa, 2010) and robust strains need to be used for genetic engineering purposes for novel biofuel production (Koffas, 2009). Importantly to note is that large scale cultivation of genetically modified strains can pose a risk of escape and contamination of the surrounding environment and crossing with native strains. Thus cultivation of engineered strains can have unintended consequences to public health and the environment (Parmar et al., 2011). There is no doubt that the solution to these challenges will need a multidisciplinary approach. For example, in order to cover all these challenges in an integrative manner, people from the fields of fundamental biology, systems biology, metabolic modeling, strain development, bioprocess engineers, scale-up and biorefinery need to work hand in hand to realize the full potential of microalgal biofuel production (Wijffels and Barbosa, 2010, Wijffels et al., 2010).

Although a wide variety of biofuels are produced routinely on the lab scale, one of the major hurdles in biofuel production is associated with the cost of its commercial production (Wijffels and Barbosa, 2010, Peralta-Yahya et al., 2012, Steen et al., 2010). Key is to find processes that will aid in reducing the overall fuel production costs. Product-recovery efficiency is an important determinant of the total production cost and a major limitation of biochemical production from photosynthetic organisms (Atsumi et al., 2009). It is estimated that 70-80% of the total cost of biofuel production comes from the biomass extraction processes as these are very energy intensive (Molina Grima et al., 2003, Robertson et al., 2011). Therefore finding means of extracting the produced fuels in a more effective manner would aid in the reduction of the overall cost of biofuel

production. It is apparent that in order to achieve such goals both the fuel-producing hosts and pathways must be engineered and optimised (Lee et al., 2008, Peralta-Yahya and Keasling, 2010). Modifying microbial pathways will require the employment of advanced technologies such as synthetic biology and metabolic engineering tools such as those described in section 1.1.4. An important challenge facing biofuel production through microbial metabolic engineering is satisfying cellular energetic concerns (thermodynamic constraints) (Fortman et al., 2008). Novel biofuel production requires re-wiring of the native pathway or the introduction of a heterologous pathway in the engineered strains, which ultimately leads to carbon flux imbalance. Therefore, it is essential to redirect the carbon flux and finely control the fixed carbon partitioning in central metabolic pathways, something that is still challenging in the field of metabolic engineering (Melis, 2013). One way of approaching this challenge is by targeting the non-growth phase of the engineered photosynthetic organisms. Since it has been demonstrated that biofuels can be produced by metabolically active cells even when they have ceased growing (Atsumi et al., 2009, Wang et al., 2013, Liu et al., 2011). However, in order for gain full advantage of this opportunity, complete genetic circuits need to be designed that remains highly active during the stationary phase (Berla et al., 2013).

Many researchers have investigated the challenges associated with biofuel production (Dismukes et al., 2008, Lu et al., 2011, Robertson et al., 2011, Chisti, 2007, Parmar et al., 2011), the following section aims at highlighting some of the main concerns.

Photobioreactor design: whether shallow big ponds, open or closed ponds, flat panels, closed columns or tubular bioreactors are used for culturing organisms for biofuels production, the reactor design should include factors that aim to optimise solar insulation and distribution (Chisti, 2007), maintain a good culture density and manage properly the gas mass transfer, mixing and temperature (Robertson et al., 2011). There is no bioreactor design that has been developed that meets all these criteria thus far. Nevertheless, a lot of efforts are currently been exerted in the development of photobioreactors that incorporate the discussed features (Parmar et al., 2011, Hankamer et al., 2007, Molina Grima et al., 2003, Cogne et al., 2005, Chisti, 2007). Figure 1.16 shows examples of different photobioreactor designs used for algal/cyanobacterial cultivation.

Downstream processing of cultures: typically microalgal cultures are very dilute and require several techniques for their harvesting. The problems lie in the high energy requirements associated with the multiple steps involved in harvesting the cells (including flocculation, sedimentation, flotation, and finally centrifugation or filtration) (Parmar et al., 2011). In addition, the cost associated with the actual lipid extraction needs to be addressed. This is due to the harsh conditions required to break the sturdy cell walls of the microalgae (Wijffels and Barbosa, 2010, Parmar et al., 2011), the cost of transesterification of the TAGs, and the significant amount of glycerol produced as a byproduct (Liu and Curtiss, 2009, Chisti, 2007). All these processes add significantly to the cost of the produced biofuel (Lu, 2010). On the other hand, the cultures are normally dewatered post harvesting in order to obtain the actual fuel from the culture. High water content and high nitrogen and phosphate contents are major limitations in downstream processing of microalgae. Furthermore, issues such as energy cost associated with downstream processing, the plant site, transportation, water quality and recycling have to also be considered for efficient biofuel production (Parmar et al., 2011). The idea of creating microalgae that can self-destruct (Liu and Curtiss, 2009) would aid in both reducing the cost and achieving quicker recovery of the product.

Photosynthetic efficiency: Although algae and cyanobacteria have better photosynthetic efficiency than terrestrial plants, culturing strategies that promote high photosynthetic conversion efficiencies are still needed for microalgae in order to maximise the photosynthetic efficiency. Employing genetic engineered cyanobacterial strains is one way of increasing the photosynthetic efficiency. The theoretical maximum photosynthetic efficiency of engineered cyanobacteria can reach a striking 10-12% (Robertson et al., 2011, Parmar et al., 2011, Ducat et al., 2011b) compared to un-engineered strains (7.2%) and algae growing in open ponds (1.5%) [Parmar et al. (2011), reports 5%]. While conventional crops such as corn or sugar cane have a photosynthetic efficiency of just 1% (Parmar et al., 2011).

Accumulated by-products & product toxicity: Product toxicity is a major concern in the production of advanced biofuels and needs to be addressed (Lu et al., 2011). For example, Atsumi et al. (2008a), mentioned that it is necessary to improve the tolerance of microorganisms for advanced biofuel production in order

to overcome the current limitations of the yield of the product. *E. coli* for example, can only tolerate up to 10 g/L of butanol if the tolerance is not increased genetically. Issues related to biofuel product toxicity are covered in more detail in Chapter 5.

Figure removed for copyright reasons

Figure 1.16: Examples of different photobioreactor designs

(A)Flat panels bioreactor, (B) Tubular low cost bioreactor, (C) Open pond bioreactors, (D) Closed tubular bioreactors,(E) An architectural design of bioreactors of the future. Images were obtained from the internet as detailed in Appendix 16.

1.4 Isoprenoids

Isoprenoids are naturally occurring branched or cyclic hydrocarbons (Lee et al., 2008) made from a branched five-carbon precursor that are synthesised in all living organisms (Ershov et al., 2002). They are considered the most ancient and diverse class of natural products (Lange et al., 2000) encompassing over 30,000 known compounds (Sacchettini and Poulter, 1997, Ershov et al., 2002) and were reported in the sediments as long as 2.5 billion years ago (Summons et al., 1999, Brocks et al., 1999).The isoprenoid pathway in plants, bacteria, microalgae and cyanobacteria has been of interest to many researchers for its essential cellular compounds (Lange et al., 2000, Lichtenthaler, 2000, Martin et al., 2003, Ro, 2006, Yoon et al., 2007, Lee and Schmidt-Dannert, 2002, Chang et al., 2007). The isoprenoid biosynthetic pathway can be exploited for the production of several branched-chain alcohols, alkanes, alkenes and cyclic hydrocarbons (Fortman et al., 2008).

1.4.1 Isoprenoid biosynthesis

There are essentially two distinct and independent biosynthetic pathways for isoprenoid biosynthesis (Figure 1.17); the mevalonate (MV) or mevalonic acid pathway (MVA), and the methyl-erythritol-4-phosphate (MEP) or also called the DXP pathway (Martin et al., 2003, Daum et al., 2009). The former is most

common in eukaryotes while the latter operates in prokaryotes (Daum et al., 2009, Lindberg et al., 2010, Dewick, 2002) including cyanobacteria (Ershov et al., 2002) and in higher plants, within the plastids (Cordoba et al., 2009, Lichtenthaler, 2000). In both pathways, Isopentenyl diphosphate (IPP) is considered the central intermediate (Lange et al., 2000) and is converted to dimethylallyl (DMAPP) via the IPP isomerase (Cordoba et al., 2009, Lindberg et al., 2010). In mammals and yeast and other eukaryotes other than plants, the pathway towards IPP synthesis starts from acetyl-CoA proceeding through the intermediate mevalonic acid (MVA) (Cordoba et al., 2009) while in the MEP pathway the synthesis starts by the condensation of pyruvate and glyceraldehyde-3-phosphate (G3P) via 1-deoxyxylulose-5-phosphate (DXP) as the first intermediate as seen in Figure 1.17 (Lange et al., 2000, Cordoba et al., 2009). This step is catalyzed by 1-deoxy-D-xylulose-5-phosphate synthase (DXS) while DXR catalyses the second step pf the pathway. Both enzymes (indicated in red) are considered rate limiting steps in the MEP pathway, as has been shown in several plants (Cordoba et al., 2009). However, examining Figure 1.17, which includes the direct link between the primary products from photosynthesis and the isoprenoid biosynthetic pathway, indicates that the G3P and pyruvate are not strictly the only substrates for the MEP pathway (sugar phosphates from the pentose phosphate cycle can also feed as substrates). Whatever the starter substrates are, the product of the two pathways leads to the production of the five-C compound Isopentenyl diphosphate (IPP) (Lange et al., 2000). For the isoprenoid synthesis, the Isopentenyl diphosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) are linked together in a head to tail condensation manner, to form the precursors geranyl pyrophosphate (GPP, C10) which is considered the universal precursor of all monoterpenes (Iijima et al., 2004, Oswald et al., 2007), farnesyl pyrophosphate (FPP, C15) and geranylgeranyl pyrophosphate (GGPP, C20), respectively. Finally, fuel ready molecules such as geraniol, farnesol and geranylgeraniol can then be synthesised from GPP, FPP or GGPP respectively by introducing a single additional enzyme (a terpene synthase) (Rude and Schirmer, 2009, Peralta-Yahya and Keasling, 2010) as seen in the same figure. The classification of the terpenoids are generally based on the number of IPP units from which they derive (Daum et al., 2009).

Figure removed for copyright reasons

Figure 1.17: Isoprenoid biosynthesis from the MEV and MEP pathways.

Abbreviations used: **G3P=** glyceraldehyde 3-phosphate; **DXP=**1-Deoxylulose-5-phosphate, **MEP=** 2-C-methyl-D-erythritol, 4-phosphate, **CDP-ME=** diphosphocytidylyl methylerythritol; **CDP-MEP=** CDP-ME 2-phosphate; **ME-cPP=**methylerythritol 2,4-cyclodiphosphate; **HMBPP=**hydroxymethylbutenyl diphosphate; **IPP=**isopentenyl diphosphate; **DMAPP=** dimethylallyl diphosphate; **HMG-CoA:=** Hydroxymethylglutaryl-CoA **.** *Enzymes*: **DXS=** DXP synthase; **DXR=** DXP reductoisomerase; **IspD=** CDP-ME synthase; **IspE=**CDP-ME kinase; **IspF=**ME-cPP synthase; **GcpE (IspG)=**HMBPP synthase; **IspH=**HMBPP reductase;**Ipi/Iidi=** IPP isomerase; **atoB=**acetoacetyl-CoA thiolase; **HMGS=**hydroxymethylglutaryl-CoA synthase; **tHMGR=**truncated hrdroxymethyl-glutaryl-CoA reductase; **Erg1=** MEV kinase; **Erg8=**phosphomevalonate kinase; **MVD1=**MEV pyrophosphate decarbonylase**; gpps=** geranyl pyrophophate synthase**; fpps=**farnesyl pyrrophosphate synthase**; ggpps=** geranyl-geranyl pyrophosphate synthase**; ts=** terpene synthase. Modified from (Peralta-Yahya and Keasling, 2010, Lindberg et al., 2010).

1.4.2 Isoprenoid biosynthesis in *Synechocystis*

The MEP pathway in cyanobacteria is vital for the synthesis of terpenoid-based molecules that are needed for essential cell functions such as photosynthesis, membrane stability and carotenoid production (Poliquin et al., 2004). *Synechocystis* appears to have genes that are homologous to the genes of the MEP pathway present in *E. coli* that leads to the synthesis of IPP and DMAPP (Figure 1.17). However, work done by Ershov et al. (2002), showed that the initial substrate of the MEP pathway in *Synechocystis* is not the same as that of *E. coli* (not pyruvate or deoxyxylulose-5-phosphate). Furthermore, he was unable to demonstrate the activity of the IPP isomerase in the cyanobacterium and therefore suggested that the interconversion of DMAPP and IPP is unlikely, which is expected due to the absence of the *ipi* gene in its genome (Kaneko et al., 1996). These results suggest that the MEP pathway as presented in *E. coli* is not the primary pathway by which isoprenoids are synthesised in *Synechocystis* under photosynthetic conditions and suggest that the products of photosynthesis (*i.e.* metabolites of the pentose phosphate pathway cycle) lead to the production of DMAPP (Figure 1.17). In this context, it is interesting to note that during photosynthesis, the majority of the carbon flux (80-85%) is directed towards sugar biosynthesis, approximately ~10% is for fatty acid biosynthesis while terpenoid biosynthesis accounts for only 3-5% (Lindberg et al., 2010). It is clear that in order to increase the titre of novel isoprenoids, metabolic engineering

strategies would need to be employed to shift the carbon flux away from sugar biosynthesis and more towards isoprenoid biosynthesis.

1.4.2 Application of isoprenoids

Because of their diversity, isoprenoids exhibit many variable biological functions, for example isoprenoids have been exploited for their pharmaceutical value (antibiotic, anticancer drugs and hormones) (Daum et al., 2009) and their nutritional value for many years (Rude and Schirmer, 2009). However, isoprenoids also serve many other functions including; quinones in electron transport chains, components in membranes, involved in subcellular targeting and regulation, as photosynthetic pigments, fragrance compounds, and as plant defense compounds (Lange et al., 2000, Martin et al., 2003). They are also important components of vitamins, essential oils, rubber, carotenoids and plant hormones (Ershov et al., 2002). Moreover, there are different structures accessible via the isoprenoid biosynthetic pathway that can be exploited for the production of biofuels as alternatives to gasoline, diesel and jet fuel (Peralta-Yahya and Keasling, 2010). It has been reported that pure hydrocarbons are generally considered better biofuels than sugars and alcohols as they store more energy in the hydrocarbon molecules (Schakel et al., 1997). It is for these reasons that isoprenoids have recently been exploited for novel biofuel production. *E. coli* platforms (Carter et al., 2003, Peralta-Yahya and Keasling, 2010) and *S. cerevisiae* platforms (Chambon et al., 1990, Chambon et al., 1991, Oswald et al., 2007) are already in place for the production of a wide range of isoprenoids. For example, the productions of the sesquiterpenes farnesol and farnesene, which are both derived from FPP, have already been achieved in both species using truncated yeast phosphatase (Song, 2006) and a plant terpene synthase (Martin et al., 2004), respectively. The C5 terpene, isoprene, has also recently been produced in cyanobacteria (Lindberg et al., 2010). However, one of the main obstacles for efficient microbial biosynthesis remains the production of the universal precursors IPP and DMAPP (Peralta-Yahya and Keasling, 2010) and the establishment of a photosynthetic platform for isoprenoid biosynthesis, something that was attempted in work presented in chapter 5.

1.5 Objectives of the presented work

This thesis describes genetic tools developed for the creation of *Synechocystis* transformants that will ultimately aid in the advancement of cyanobacterial biofuel research.

The aims of the work undertaken are:

- To develop a quick method to generate random *Synechocystis* mutants using a gene tagging approach based on the random insertion into the genome of a marker gene. This is covered in chapter 3.
- To develop a rapid one-step PCR based method for creating transforming DNA for targeted gene deletions in *Synechocystis* that will result in eliminating the need to construct plasmids for transformation. This is the topic of chapter 4.
- To create a photosynthetic "cell factory" for the synthesis of selected isoprenoid-derived fuels, namely geraniol and farnesene, and assess their toxicity on the cyanobacterial host. Strain creation, biofuel production and toxicity tests are covered in chapter 5.
- To test the function of the AlkL membrane protein as a hydrocarbon pump in *Synechocystis* and compare it to its recently established role in *E. coli.* Chapter 6 covers work related to AlkL, which was carried out in collaboration with Dr Christopher Grant from the Department of Biochemical Engineering at University College London.

Chapter 2: Materials and Methods

2.1 Chemicals

The chemicals used in this work were of the highest chemical grade available. Unless otherwise specified, the chemicals were purchased from Sigma Aldrich and all solutions were prepared using double distilled water (ddH_2O) from a Maxima water filtration unit (Elga, High Wycombe). Solutions that required sterilisation were autoclaved (Priorclave Midas 56, UK) at 121^0C and 1 atm for 15 min and handled using aseptic techniques. All restriction enzymes used in this work were purchased from New England Biolabs (NEB).

2.2 Strains employed, their growth and storage conditions

2.2.1 *Escherichia coli* bacterial strains

The bacterial strain used was *Escherichia coli* DH5α, {genotype: F-, (φ80lacZΔM15), Δ(lacZYA-argF)U169, deoR, recA1, endA1, hsdR17(rk-, mk+), supE44, thi-1, gyrA96, relA, λ} supplied by CLONTECH Laboratories, Inc. This stain will be referred to as *E. coli* hereafter.

2.2.2 Growth and storage of *Escherichia coli*

E. coli was grown in LB (lysogeny broth) medium containing 10 g/L tryptone (Difco), 5 g/L bacto yeast extract (Difco), and 10 g/L NaCl. For LB plates, 15 g/L Agar was added to the LB medium (Bertani, 1951). Cultures were grown at 37^oC in shacking incubators [Stuart Scientific Incubator S.I.60, UK with built-in shaker IKA-VIBRAX-VXR, IKA® Labortechnik, Germany] at 170 rpm. The LB plates were incubated overnight in a static incubator [LAB. Companion, Fisher Scientific].

For short-term storage the bacteria grown in liquid or solid media were stored at 4^oC. Where appropriate, the autoclaved and cooled medium (< 50^oC) was supplemented with ampicillin, kanamycin or zeocine at concentrations of 75 µg/ml, 50 µg/ml and 25 µg/ml, respectively.

For long-term storage, *E. coli* strains were stored as 20% (v/v) glycerol stocks. This was achieved by mixing equal volumes (~0.5 ml) of grown bacterial cultures with 80% (v/v) autoclaved glycerol, vortexed briefly, snap frozen in liquid nitrogen and stored at -70^oC or below (Williams, 1988).

2.2.3 *Synechocystis* strains and mutants created

The wild type (WT) (large, glucose tolerant) *Synechocystis* strain was obtained from Prof Conrad Mullineaux, Queen Mary University of London.

The *recJ* deleted mutants of *Synechocystis* were created by deleting the *recJ* gene (gene code: sll1354) in the wild type *Synechocystis* strain. This was done by inserting either the *ble* cassette or kanamycin resistance (km^R) gene in the *recJ* gene and transforming the corresponding plasmids carrying the deleted genes into WT *Synechocystis* creating the Δ*recJ*.ble and the Δ*recJ*.km^R *Synechocystis* mutants, respectively. Details of the plasmids created to make the mutants can be found in Appendix 1.

2.2.4 Growth and storage of *Synechocystis* strains

Synechocystis strains were grown photoautotrophically on solid BG11 medium prepared by mixing equal volumes of 2X BG11 medium with 3% (w/v) agar after autoclaving (Table 2.1). Cultures on agar plates were stored in a lit incubator at 30^oC with a light intensity of 30-50 µmol/m^2/s for two to three weeks (Castenholz, 1988). Single colonies were re-streaked every 3-4 weeks on fresh BG11 plates and stored under the same conditions to confirm the growth and isolation of axenic *Synechocystis* strains. Cells were grown in liquid culture at 30^oC at 150 rpm under a light intensity of 20-50 µmol/m^2/s. All microbiological manipulations were done in a laminar flow hood.

All mutants of *Synechocystis* were grown under the same culture conditions. Where appropriate, BG11 plates were supplemented with either kanamycin or zeocin at concentrations of 200 µg/ml and 25 µg/ml, respectively.

For long-term storage, cells were stored in 15% glycerol. Cells from 100 ml of liquid culture (OD_{750}: 1-2) were harvested by centrifugation (MISTRAL1000, MES Scientific Instruments, Loughborough) at 4,500 x *g* for 6 min at room temperature. Cell pellets were re-suspended in residual growth medium present in the pellet; the volume of the dense suspension was measured, then mixed with 0.24 volumes of autoclaved 80% glycerol before being stored at -80^oC (Bishop, 2002). For recovery, the cells were thawed and either spread onto plates or inoculated into liquid medium.

Table 2.1: Recipe of BG11 medium for *Synechocystis*

1 L 1X BG11 liquid medium	Amount
100X BG11 [(a)]	10 ml
Trace elements [(b)]	1 ml
Iron stock [(c)]	1 ml
ddH_2O	to 1 L
Autoclave and allow to cool then add	
Phosphate stock [(1)]	1 ml
Na_2CO_3 stock [(2)]	1 ml
$NaHCO_3$ stock [(3)]	10 ml
TES buffer [(4)]	10ml

(Castenholz, 1988)

For BG11 plates, 2X concentrated BG11 was prepared including 3 g of sodium thiosulphate and mixed with Difco Bacto-agar (3% w/v in distilled water) to give 1 L of BG11 plates (1X BG11 and 1.5% agar). Details for the preparation of solutions a, b, c and 1,2,3 and 4 can be found in Appendix 2.

2.2.5 Quantification of *Synechocystis* cultures

The optical density (OD_{750}) of the liquid culture was measured at 750 nm using a Unicam UV-Vis spectrometer (Thermo Electron Corporation, Cambridge). The concentration of cells (cells/ml) was calculated by multiplying the figure obtained by 1.15×10^8 (Bishop, 2002).

2.2.6 Growth tests on solid media "spot tests"

Spot tests of strains of interest were performed where small aliquots of the cultures (~6 µl) at the mid-log phase were applied on BG11 plates supplemented with the appropriate antibiotic when necessary. The spots were allowed to dry completely before incubating them at 30^oC with a light intensity of 30-40 µmol/m^2/s for 5-7 days depending on the tested strain. Wild type strains were always included as controls.

2.3 DNA techniques

2.3.1 Expression vectors

The list of expression vectors used in this study can be found in Table 2.2

Table 2.2: Plasmids, their source and usage.

Expression vector	Source	Usage
pUC.psbN.km^R.A	Saul Purton's collection	pUC9-based vector for targeted incorporation of a km-resistance cassette into a neutral locus in *Synechocystis* (*psbN*)
pUC4K	(Taylor and Rose, 1988)	source of a km-resistance gene
pzΔES	Saul Purton's collection	source of a ble cassette
pUC18	(Yanisch-Perron et al., 1985)	base for creating the two expression vectors pLAH.A2 and pLAH.nrsB (section 2.3.11)
pskΔE	Saul Purton's collection	a pBluescript.sk based plasmid used in the creation of the Δ*recJ*.ble and subsequently the Δ*recJ*.km plasmids
pCRT7/CT-TOPO	Eran Pichersky (Iijima, Gang et al. 2004)	source of the geraniol synthase gene (GES) (sequence of the CDS of GES can be found in Appendix 4)

Classical molecular techniques were carried out for DNA manipulation as described in (Sambrook and Russell, 2001).
Appendix 1 details the plasmids utilised and created in this work.

2.3.2 Plasmid DNA isolation

Small scale plasmid DNA isolations were obtained using 1.5-3 ml of a 10 ml bacterial culture using a QIAprep Spin Miniprep Kit (QIAGEN, Crawley, UK) according to the manufacturer's instructions. This was used mainly to isolate plasmids for transformation and sequencing purposes where clean preparations were essential.
For crude plasmid extracts, a simple alkaline lysis method was used (Birnboim and Doly, 1979). In this method cells from 3-5 ml of an overnight grown *E. coli* culture were spun down, the supernatant discarded and 250 µl of re-suspension buffer (50 mM Tris pH 8.0 with HCl, 10mM EDTA, 100 µg/ml RNase A) was added to the pellet. The cells were lysed by adding 250 µl of lysis buffer (200 mM NaOH, 1% SDS, freshly prepared) followed by the addition of 350 µl of neutralising solution (3.0 M potassium acetate, pH5.5) and mixing the solution by inverting it several times. The mixture was centrifuged for 10 min at 16,200 x *g*

(Biofuge Pico, Heraeus) and the lysate transferred to a fresh 1.5 ml eppendorf tube. Absolute ethanol was added to precipitate the plasmid DNA followed by a 15 min spin that resulted in the precipitation of a glossy pellet. The pellet was washed with 70% ethanol and allowed to air dry before the plasmid DNA was re-suspended in 50 µl of ddH_2O.

2.3.3 *Synechocystis* genomic DNA isolation

2.3.3.1 Genomic DNA isolation for PCR

The Chelex-based method adapted from Werner and Mergenhagen. (1998), was used for the extraction of genomic *Synechocystis* DNA. In this method, a loopful of cells from a BG11 plate was re-suspended in 20 µl of sterile ddH_2O. 20 µl of absolute ethanol was added and the mixture was incubated at room temperature for 1 min. 200 µl of a 5% v/v suspension of Chelex 100 resin (Bio-Rad) was added and the samples were mixed briefly using a vortex and placed in a boiling water bath for 5 min. The samples were then centrifuged at 16,200 x *g* for 2 min (Biofuge Pico, Heraeus) after which the supernatant was transferred to a fresh tube. 2 µl of this extracted DNA was used in 50 µl PCR reactions (section 2.3.5).

2.3.3.2 Total genomic DNA for southern blot analysis

Genomic DNA was isolated from 20 ml of stationary phase *Synechocystis* cells (OD_{750} of ≥1.0). The cells were harvested by centrifugation in a bench top centrifuge (MISTRAL1000, MES Scientific Instruments, Loughborough) at 3000 x *g* for 15 min. The supernatant was discarded and the pellet re-suspended in 400 µl TES buffer (composed of 5 mM Tris.Cl pH 8.5, 50 mM NaCl and 5 mM EDTA). It was then transferred to a 2 ml Eppendorf tube where the cells were lysed by the addition of lysozyme (10 mg/ml) followed by incubation for 45 min at 37^{o}C. Next, 50 µl sarkosyl and 600 µl saturated phenol pH 8.0 (BDH, Germany) were added to the tube. The tubes were mixing using a vortexer for 10 min after which, 100 µl of 5 M NaCl, 100 µl CTAB-NaCl (700 mM NaCl, 10% (w/v) CTAB) and 600 µl chloroform:isoamylalcohol (24:1 v/v) were added. The tubes were mixed again in the same manner (once a minute for 10 min). Centrifugation on a bench top micro-centrifuge (HERAEUS PICO21, Thermo Scientific, Germany) at 1500 x *g* for 2 min produced a nucleic acid-containing top phase which was transferred to a fresh 2 ml Eppendorf tube where a second phenol-chloroform extraction was

performed using equal volumes of saturated phenol pH8.0 and chloroform:isoamylalcohol (24:1 v/v).

Precipitation of the genomic DNA was achieved by the addition of 1 ml of cold ethanol, followed by 20 min incubation at -20^{o}C and micro-centrifugation in a bench-top centrifuge (HERAEUS PICO21, Thermo Scientific, Germany) at 3000 x *g* for 10 min. The pellet was washed with 70% (v/v) ethanol for 5 min at room temperature then dried at 37^{o}C for 5 min. The DNA was finally dissolved in 50 µl of ddH_2O/RNaseA (10 µg/ml) (Bishop, 2002).

This procedure produces enough DNA of sufficient purity for two to four restriction digests for Southern analysis.

2.3.4 Preparation of competent *E. coli*

To create competent *E. coli* cells, the DH5α strain (2.2.1) was re-streaked from frozen glycerol stocks on a fresh LB plate and incubated overnight at 37^{o}C. A single colony was used for inoculation of 5 ml LB medium and incubated under the same conditions. 1 ml of the grown culture was used to inoculate a 100 ml of fresh LB medium and allowed to grow for 2.5 hours at 37^{o}C and shaking at 200 rpm. The 100 ml culture was then placed on ice for 15 min, pelleted in pre-cooled 20 ml tubes [Sterilin Ltd, Thermo Scientific, UK] at 3000 x *g* for 5 min. The cell pellet was re-suspended in 40 ml of ice-cold 50 mM $CaCl_2$ and kept on ice for 30 min. The cell suspension was pelleted another time under the same conditions and re-suspended once more with fresh ice-cold 50 mM $CaCl_2$ to up to 8 ml. The solution was mixed with 3.5 ml of 50% (v/v) glycerol, aliquots of 200 µl were prepared and frozen in liquid nitrogen and stored at -80^{o}C.

2.3.5 Polymerase chain reaction

For the amplification of either plasmid or genomic DNA a programmed thermocycler was used (Eppendorf mastercycler® personal, Hamburg) or (Techne TC-3000X Thermal Cycler, California, USA). Reactions were performed in a total volume of 50 µl and the reaction mixture comprised: 10 µl 5X Phusion HF buffer (New England Biolabs), 1 µl of 20 mM mixed dNTPs (Fermentas, York, UK), 0.5 µl of forward and reverse Eurofins MWG Operon primers (Ebersberg, Germany) with a final concentration of 100 pmol/µl, 0.5 µl (1U) of Phusion® High-Fidelity DNA Polymerase (New England Biolabs), 1 µl template DNA

(aproximately 1 µg genomic or 10 ng plasmid DNA). Finally, ddH_2O was added to the reaction mixture to make a final volume of 50 µl.

For PCR reactions using Taq polymerase (NEB), reactions were performed in 25 µl reaction mix where 0.75 µl of 25 mM $MgCl_2$ was added to 2.5 µl of the 10X PCR buffer supplied with the enzyme. Then, 0.5 µl of 10mM dNTPs, 1 µl of template DNA and 0.25 µl of each primer and 1 µl of Taq polymerase (NEB) were used and the final volume was reached by adding ~19 µl of ddH_2O.

For GC rich DNA, 1.5 µl 100% DMSO was added to the reaction mixture and 5X Phusion GC rich reaction buffer was used.

The steps were carried out according to the manufacturer's instructions and where necessary; the programme included a 'hot start' where the DNA template was incubated at 98^{o}C for 1 min before adding the Phusion® High-Fidelity DNA Polymerase to the reaction.

2.3.6 PCR purification and sequencing

PCR products were purified using the QIAquick PCR purification Kit (QIAGEN) according to the manufacturer's instructions. DNA samples that required sequencing were sent to the scientific support services at the Wolfson Institute for Biomedical Research at University College London. The samples were sequenced using BigDye3.1 chemistry (Applied Biosystems) and analysed using an Applied Biosystems AB3730XL genetic analyser. The returned data was checked manually for base-calling errors using Emboss align [4] and NCBI thermo Align [5].

2.3.7 Oligonucleotide primers

Appendix 3 details the primers used in this work and the experiments in which they were used. All primers were synthesised by Eurofins MWG (Ebersberg, Germany).

---

[4] http://www.ebi.ac.uk/Tools/psa/

[5] http://blast.ncbi.nlm.nih.gov/Blast.cgi

2.3.8 Codon optimisation

For the genes in this study that required synthetic production, GeneArt (www.geneart.com) was consulted where they carried out both gene codon optimisation and gene synthesis.

2.3.9 Agarose gel electrophoresis

DNA fragments were separated according to size on 1% (w/v) agarose gels made in 1X TAE buffer (40 mM Tris, 1 mM sodium EDTA, 17.5 mM glacial acetic acid). To run the gel, 6X DNA Loading Dye & SDS solution (Fermentas) was added to the DNA sample and the gel was pre-stained with ethidium bromide (1 µg/ml). The gels were submerged in 1 x TAE buffer in an electrophoresis tank (Sigma) and allowed to run for 90 min at 85 V (Gibco power supply). Finally, the gels were illuminated on a UV transilluminator (UVP Gel Documentation System) to visualise the bands. O'GeneRuler™ 1 kb Plus DNA Ladder, ready-to-use (Fermentas) was used to estimate the size of the DNA fragments.

2.3.10 Recovery of DNA from agarose gels

To recover the DNA from the gel, the gel was placed on a UV plate (UVP, California) and the area that contained the DNA of interest was excised using a razor blade. The DNA was recovered using the QIAquick Gel Purification Kit (GIAGEN, Crawley) according to the manufacturer's instructions.

2.3.11 Construction of *Synechocystis* expression vectors

2.3.11.1 Expression vectors

For heterologous expression of foreign genes in *Synechocystis*, plasmid constructs were created to allow the replacement of the dispensable *psbA2* coding sequence (gene code = slr1311) with the CDS of the gene of interest via double homologous recombination. This is covered in detail in chapter 5 section (5.3.1).

2.3.11.2 Restriction digests and cloning

1 µg of DNA was cut with at least 1 unit of restriction enzyme in the recommended buffer (New England Biolabs, Hitchin). Dilutions of the digestion were done using ddH_2O, and supplemented with 1X BSA when necessary. All steps were carried out according to the manufacturers' instructions. The pJET

vector (Clone JET PCR Cloning Kit, Fermentas, Germany) was used for routine cloning according to the manufacturer's instructions.

2.3.11.3 DNA Ligation

Linearised vector, insert DNA, T4 DNA ligase buffer and T4 DNA ligase were added according to the manufacturer's instructions to give a total volume of either 10 µl or 20 µl. The ratio of insert: vector DNA was ~ 6:1. Ligation reactions were typically incubated for ≥ 2 hours at room temperature. The ligation mix was used directly for transformation of the competent DH5α strain of *E. coli*.

2.4 Transformation technique

2.4.1 Transformation of competent *E. coli*

Competent DH5α cells (section 2.3.4) were thawed and placed on ice. 1 µl of a plasmid mini preparations (diluted 1/10) or 10-20 µl of the ligation mix (section 2.3.11.3) were added to the 100 µl thawed competent cells and the mixture was incubated on ice for at least 30 min before subjecting it to a heat shock at 42^{o}C for 60 seconds. It was then placed briefly on ice followed by the addition of 1 ml of fresh sterile LB medium. The cell suspension was incubated at 37^{o}C for 1 hour in a shaking incubator (Stuart Scientific Incubator S.I.60, UK with a built-in shaker IKA-VIBRAX-VXR, IKA® Labortechnik, Germany) at 170 rpm. The cells were pelleted by centrifugation in a bench top centrifuge (HERAEUS PICO21, Thermo Scientific, Germany) at 16,200 x *g* for 2 min. The supernatant was discarded and the pellet re-suspended in 300 µl of fresh LB medium. Aliquots of 100 µl of the culture was plated on selective LB plates and incubated overnight at 37^{o}C. The selective LB plates contained ampicillin, kanamycin or zeocin at concentrations of 75 µg/ml, 50 µg/ml and 25 µg/ml, respectively depending on the antibiotic resistance marker present in the introduced plasmid. Single colonies were picked for inoculation in 5 ml of LB medium containing the suitable antibiotic for selection and incubated overnight with shaking at 170 rpm.

2.4.2 Transformation of *Synechocystis*

For the transformation of *Synechocystis*, a loopful of cells from freshly re-streaked plates was inoculated in 100 ml of liquid BG11 medium and allowed to grow in a light shaking incubator at 25^{o}C and a light intensity of 15 µmol/m^{2}/s for four days. The cultures were quantified by measuring the OD of the liquid culture

at 750 nm using a UNICAM UV/Vis Spectrometer. The concentration (cells/ml) was calculated using the method described in (2.2.5).

The transformation method employed was adapted from (Williams, 1988). Wild type or mutant *Synechocystis* with an OD_{750} between ~0.4-0.8 was harvested by centrifugation using a bench top centrifuge at 3000 x *g* for 10 min. The cells were washed in 2 ml of fresh BG11 medium and spun down a second time. The washed cell pellets were re-suspended in BG11 medium to a final concentration of 4 X 10^8 cells/µl. In duplicates, 1-5 µg of the DNA to be transformed was added to 200 µl of the re-suspended cells. These were then incubated at 30^{o}C for 4-6 hours. The cells were spread onto BG11 plates, allowed to dry, and then incubated for 2-3 days to allow expression of the introduced resistance marker.

The plates were then overlain with 3 ml of 0.6% (w/v) agar with added kanamycin or zeocin giving a final concentration of 200 µg/ml and 25 µg/ml, respectively. The plates were incubated for approximately two weeks in a light intensity of ~40 µmol/m^2/s at 30^{o}C and were then examined for the presence of putative transformant colonies.

2.4.2.1 Isolation and analysis of homoplasmic *Synechocystis* strains

Colonies were picked using sterile cocktail sticks and streaked to form single colonies on BG11 plates supplemented with the appropriate antibiotic. Single colonies were taken through three rounds of selection and homoplasmicity confirmed by PCR (2.3.5) using wild type or the untransformed strain as controls.

2.5 DNA analysis

2.5.1 Southern blotting

The methodology employed was adapted from that established by Southern (1975) and described in Sambrook and Russell. (2001). In this method, the DNA of interest was digested with the appropriate enzyme, run on a 1% agarose gel and soaked, with agitation, in denaturing solution (1.5 M NaCl, 0.5 M NaOH) for ≥ 30 min. The gel was then briefly rinsed with distilled water and soaked in neutralising solution (1 M TrisHCl pH 8.0, 1.5 M NaCl) for 20 min followed by further 20 min soak in fresh solution. The DNA was irreversibly transferred to an uncharged nylon membrane (Hybond-N, GE Healthcare) by capillary action following an overnight transfer at room temperature according to the protocol described in (Sambrook and Russell, 2001). The DNA was fixed to the

membrane by baking it for 1 hour at 80^{o}C between 3MM Whatman paper (Whatman, Maidstone).

2.5.2 Probe labelling and detection

The probe was produced using ~1 µg of template DNA produced from either a PCR product or gel purified DNA from a restriction digest. DNA labelling and detection was done using the DIG High Prime DNA Labelling and Detection Starter Kit II (Roche Diagnostics GmbH, Mannheim, Germany) following the manufacturer's instructions.

2.6 Protein analysis

2.6.1 Isolation of crude whole cell protein extracts of *Synechocystis*

Synechocystis strains of interest were re-streaked several times on BG11 plates before inoculating them in 50 ml flasks containing 25 ml of BG11 medium and allowed to grow to mid-log phase. The optical density of the cells was measured using the UV-Visible spectrometer (Thermo Electron Corporation, USA) at 750 nm. 20 ml of the culture was harvested in a bench top centrifuge at 3000 x *g* for 10 min and re-suspended in X ml of solution A (0.8 M TrisHCl pH 8.3, 0.2 M sorbitol, 1% β-mercatoethanol) where X is the reading at 750 nm. This was done to obtain samples with approximately the same cell concentration. The samples were stored at -80^{o}C.

2.6.2 Protein gel electrophoresis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out using the mini-PROTEAN* System (BioRad). Laemmli's recipe (Laemmli, 1970) was used to prepare the gels as detailed in Table 2.3 and mixed in the order listed with TEMED added last. The gel comprised a 15% resolving gel that was allowed enough time to fully polymerise (≥ 30 min). Prior to polymerisation a thin layer of ethanol was overlain on the resolving gel to ensure a flat interface between the resolving and stacking gels. This thin layer was removed by filter paper before pouring the 3.75% stacking gel and a 10-well comb was added to the gel before it started to polymerise.

Table 2.3: Laemmli gel recipe for SDS-PAGE (Laemmli, 1970).

Reagents	Volume for 2 gels
15% resolving gel: (~10 ml)	
• Acrylamide/bisacrylamide 40% stock, 37:1 [Sigma, Dorset, UK]	3.75 ml
• Resolving gel buffer (8 x stock) 3 M Tris-HCl, pH 8.8	1.25 ml
• 10% SDS	0.1 ml
• ddH_2O	4.5 ml
• 10% ammonium persulfate (AMPS) [Sigma, Dorset, UK]	0.375 ml
• *N*,*N*,*N*′,*N*′-Tetramethylethylenediamine (TEMED) [Sigma, Dorset, UK]	3.75 µl
3.75% stacking gel: (~5 ml)	
• Acrylamide/bisacrylamide 40% stock, 37:1 [Sigma, Dorset, UK]	0.47 ml
• Stacking gel buffer (4 x stock) 0.5 M Tris-HCl, pH 6.8	1.25 ml
• 10% SDS	0.05 ml
• ddH_2O	3 ml
• 10% ammonium persulphate (AMPS) [Sigma, Dorset, UK]	0.25 ml
• *N*,*N*,*N*′,*N*′-Tetramethylethylenediamine (TEMED) [Sigma, Dorset, UK]	3.75 µl

The polymerised gel was assembled into the Mini-PROTEIN* System (BioRad) and the tank filled with Tris-glycine-SDS electrophoresis buffer (0.025 M Tris, 0.192 M glycine and 0.1% (w/v) SDS, pH 8.3).

The crude protein samples were prepared by adding 11 µl of 10% SDS to 100 µl of the protein sample and boiled at 99^{o}C for 1 min in a heat block (Thermomixer Comfort, Eppendorf). The samples were spun down to remove cell debris and 30 µl of each sample were loaded on the gel. 5µl of PageRuler^{TM} prestained protein markers (Fermentas, York, UK) was used to estimate protein sizes. For good protein separation, gels were typically run for 2 hours at 120 V using a PowerPac 300 (Biorad, Hemel Hempstead, UK).

2.6.3 Western blot analysis

Following SDS-PAGE to fractionate the proteins according to size, the gel was soaked in Towbin buffer (25mM Tris, 192 mM glycine and 20% (v/v) methanol) for 30 min together with a piece of Hybond^{TM} ECL nitrocellulose membrane (Amersham GE Healthcare) and eight pieces of 3 MM Whatman paper that were all pre-cut to the gel size.

Protein transfer to the membrane was carried out using a Semi-Dry Transfer Cell TRANS-BLOT*SD (BioRad). The setup was as follows, starting at the base of the electroblotter: four pieces of 3 MM Whatman paper, the Hybond^{TM} ECL nitrocellulose membrane, the gel, and the other four pieces of 3 MM Whatman paper. A voltage of 20V was applied for 1 hour using a Fisons FEC 570 Powerpac.

2.6.4 Immuno-detection

Once the transfer was complete, the membrane was incubated overnight at 4ºC with gentle agitation in blocking solution comprised of 0.5 % (w/v) of low fat powdered skimmed milk prepared in TBS (20 mM Tris base, 137 mM NaCl, 1 M HCl pH 7.4).

Subsequently, the membrane was washed vigorously (80 rev/min) with 1X TBS-T (20 mM Tris base, 137 mM NaCl, 1 M HCl pH 7.4, 0.1% Tween-20) once for 15 min then twice for 5 min each. The membrane was then incubated at room temperature for a minimum of 1 hour on a Stuart Scientific (Stone, UK) 3D rocking platform with the primary antibody diluted in blocking solution. The membrane was then washed with 1X TBS-T for at least 45 min (a 15 min wash, then 3 X 10 min washes) followed by incubation with the secondary antibody also diluted in blocking solution. When necessary, in case of a light sensitive secondary antibody, the membrane was covered with a black cloth. Finally the membrane was rinsed with TBS-T (1 wash of 15 min followed by 3 X 5 min washes). A final wash with TBS was done if the Odyssey system was used for detection. Dilutions used for primary and secondary antibodies are listed in Table 2.4.

Table 2.4: Details of primary and secondary antibodies.

Table includes the name, source, type and the dilutions of the antibodies used in this application.

Protein	Antibody	Source	Type	Dilution factor
	Primary antibody			
HA	Anti-HA	Sigma Aldrich	Rabbit monoclonal	1:2000
Ble	Anti-Ble	CAYLA	Rabbit polyclonal	1: 2000
	Secondary antibody			
Rabbit IgG	ECL anti-rabbit IgG	Sigma (A3687)	horseradish peroxidase conjugated (from donkey)	1:10,000
	Anti-rabbit IgG (H+L)	Thermo Scientific	Dylight™ 800 conjugated (from goat)	1: 20,000
Rabbit IgG	ECL anti-rabbit IgG	Sigma (A3687)	horseradish peroxidase conjugated (from donkey)	1: 5000
	Anti-rabbit IgG (H+L)	ThermoScientific	Dylight™ 800 conjugated (from goat)	1: 15,000

To detect the antibody complex, an enhanced chemiluminescence (ECL) detection kit was used (SuperSignal (R) West Pico Chemiluminescent Substrate, supplied by Thermo SCIENTIFIC, Rockford, USA). After removing excess wash buffer, equal amounts of detection solution one and two were applied to the membrane ensuring full membrane coverage. The membrane was incubated for 1 min at room temperature and excess solution wiped off. The membrane was sealed between polythene sheets using a heat sealer and placed in a developing cassette. In the dark room, the membrane was exposed to a sheet of Hyperfilm ECL (GE Healthcare) and developed using an automatic film processor (Xograph Compact X4 film developer). The signal strength was adjusted by exposing the membrane for various times (5 sec to up to 45 min) to achieve the desired signal strength.

When protein intensity were to be compared, the Odyssey®infrared imagining system (Licor-Bioscience, Lincoln, Nebraska, USA) was used according to the manufacturer's instructions.

2.7 *Synechocystis* toxicity tests

Geraniol, geranial, linalool, farnesene and citronellene were used to assess the level of toxicity of selected terpenes to *Synechocystis*.

2.7.1 Testing for growth of *Synechocystis* in the presence of terpenes

Synechocystis was grown in a liquid culture to an OD_{750} of ~1.7. Inocula of 0.5 ml of the culture were added to each of 6 flasks containing 25 ml of growth medium (BG11). Aliquots of each terpene were added to individual flasks to a final concentration of 0.02%, 0.1%, 0.2%, 0.4% or 1% (v/v). Cultures with no added test compound were monitored to check viability during the course of the experiment. Photographs were taken every 24 hours for four days to produce a color series of culture viability.

2.7.2 Tolerance of established cultures of *Synechocystis* to terpenes

Synechocystis was cultured in shaker flasks in 25 ml of BG11 medium (Section 2.2.4). Aliquots of each of the test compounds were added to individual flasks containing established cultures of *Synechocystis* to a final concentration of 0.02%, 0.1%, 0.2%, 0.4% or 1% (v/v). The additions were done when cell growth

was in the exponential phase and the population was approximately 7 x 10^7 cells/ml. Cultures with no added test compound were monitored to check viability during the course of the experiment. Photographs were taken every 24 hours for four days to produce a visual series of culture viability.

2.7.3 Effect of a two-phase culture system on tolerance levels

Cultures (25 ml) were grown to a cell population of approximately 7 x 10^7 cells/ml before adding 0%, 0.02%, 0.04%, 0.08%, 0.1%, 0.2% and 1% v/v of the terpene followed by 5 ml of n-dodecane (Alfa Aesar) to create an organic phase above the aqueous medium. Cultures with 0–1% v/v test compound were also grown in the absence of n-dodecane for comparison. Photographs of the cultures were taken every 24 hours for four days.

2.8 Biofuel and hydrocarbon analysis

2.8.1 Supernatant extracts from cyanobacteria cells

Cyanobacteria strains expressing α-farnesene synthase under the control of the *nrsB* inducible promoter were used for this extraction method. 450 ml of un-induced culture grown to mid-log phase was centrifuged for 15 min at 10,722 x *g* using a Sorval Evolution RC (Kendro Lab Products, USA). The supernatant was discarded and the pellet re-suspended in 60 ml of fresh BG11. Of the 60 ml of re-suspended cells, 15 ml were harvested in falcon tubes by centrifugation in a bench top centrifuge at 3000 x *g* for 2 min. The supernatant was transferred into a new falcon tube following the addition of 3 ml of hexane while the pellet was later used for cell extraction (Section 2.8.2). The supernatant/hexane mix was vigorously mixed using a vortex and the phases allowed to separate before transferring the top phase into gas chromatography (GC) vials for GC analysis. This was used as a negative control. The remaining 45 ml culture was induced with $NiCl_2.6H_2O_4$ added to a final concentration of 6.4 µM. The culture was incubated at 30^oC under a light intensity of 30 µmol/m^2/s for 4 hours then later split into three 15 ml cultures. One of the 15 ml induced cultures was harvested with hexane and the pellet saved as described above while the remaining two cultures were spun down, their supernatant discarded and the pellets used for cell extraction (Section 2.8.2). All flasks used were acid washed with 5 M HNO_3 overnight.

2.8.2 Cell extracts

Cell extracts for biofuel analysis: In this method, three high-performance liquid chromatography (HPLC) grade solvents (hexane, methanol and ethyl acetate) were used for farnesene extraction from the pellets. The pellet in one falcon tube (section 2.8.1) was re-suspended in 1 ml of hexane and transferred to a 2 ml Eppendorf tube containing ~500 µl of acid washed 212-300 µM glass beads (Sigma). The same process was carried out using 1 ml of ethyl acetate and 1 ml of methanol. The tubes were mixed vigorously using a vortexer for 2 min followed by 2 min incubations on ice. The vortex/ice incubation steps were performed 5 times before spinning the cells for 5 min at 16,200 x *g* (Biofuge Pico, Heraeus) after which, the top solvent phase was transferred into GC vials for GC analysis. An un-induced culture of the farnesene synthase expressing strain was used as a negative control. The samples were run in parallel with standards of commercial β-farnesene prepared in hexane at concentrations of 1 ppm, 50 ppm and 100 ppm that also served as positive controls.

Cell extracts for hydrocarbon analysis: In order to determine the hydrocarbons in the microwells, the contents of the cell were removed 21 hours post dodecane addition and transferred to pre-weighed 15 ml polypropylene falcon tubes for extraction in ethyl acetate directly. They were then centrifuged in a bench top centrifuge (HERAEUS PICO21, Thermo Scientific, Germany) at 16,200 x *g* for 15 min after which the supernatant was tipped into another 15 ml falcon tube and pellets were spun down once more for 4 mins. The remaining supernatant was carefully transferred to the falcon containing the supernatant where 2 ml ethyl acetate was added to them. They were vortexed and left in the heated cabinet for 30 mins after which they were re-vortex and spun down for 4 mins at 16,200 x *g*. 1 ml of the top phase was transferred into the GC vials for analysis. The extraction method using ethyl acetate was done according to (Grant et al., 2012). For pellet extractions, 1 ml ethyl acetate was added to the pellet, and it was vigorously shaken and placed in a heated cabinet for 30 mins. Two pre-weighed control tubes were included that only contained 1 ml ethyl acetate. This was done to determine the effect of adding ethyl acetate on the mass of tube (effect on dry cell weight measurements). The tubes were vortex, centrifuged for 4 mins at 16,200 x *g* and finally the liquid was transferred to vials for GC analysis.

2.8.3 Two-phase extraction method

In this method, *Synechocystis* strains expressing α-farnesene synthase under the *psbA2* and *nrsB* promoters (strain one and strain two respectively) and a negative control (Wild type -WT- *Synechocystis* strain transformed with an empty pLAH.A2 vector) were analysed. 300 ml of each strain were grown in acid washed flasks to an OD_{750} of ~0.5, at which, strain two was induced with 6.4 µM of $NiCl_2.6H_2O_4$ for 4.5 hours. Two non-toxic solvents namely, n-dodecane and decane were used for terpene extraction. The cultures were overlaid with 5 ml of the solvent tested and placed in a shaking incubator set at 120 rpm. They were allowed to shake for 17 hours to allow enough time for the terpene to partition into the solvent layer. The solvent phase of the two-phase culture system was collected and centrifuged for 10 min at 4000 x *g* to remove cell debris, and subsequently subjected to GC-MS. Samples spiked with 0.02% and 0.2% (v/v) terpene standard were used to check the extraction method using each solvent. The sample set up and the extraction method is illustrated in Figure 2.1.

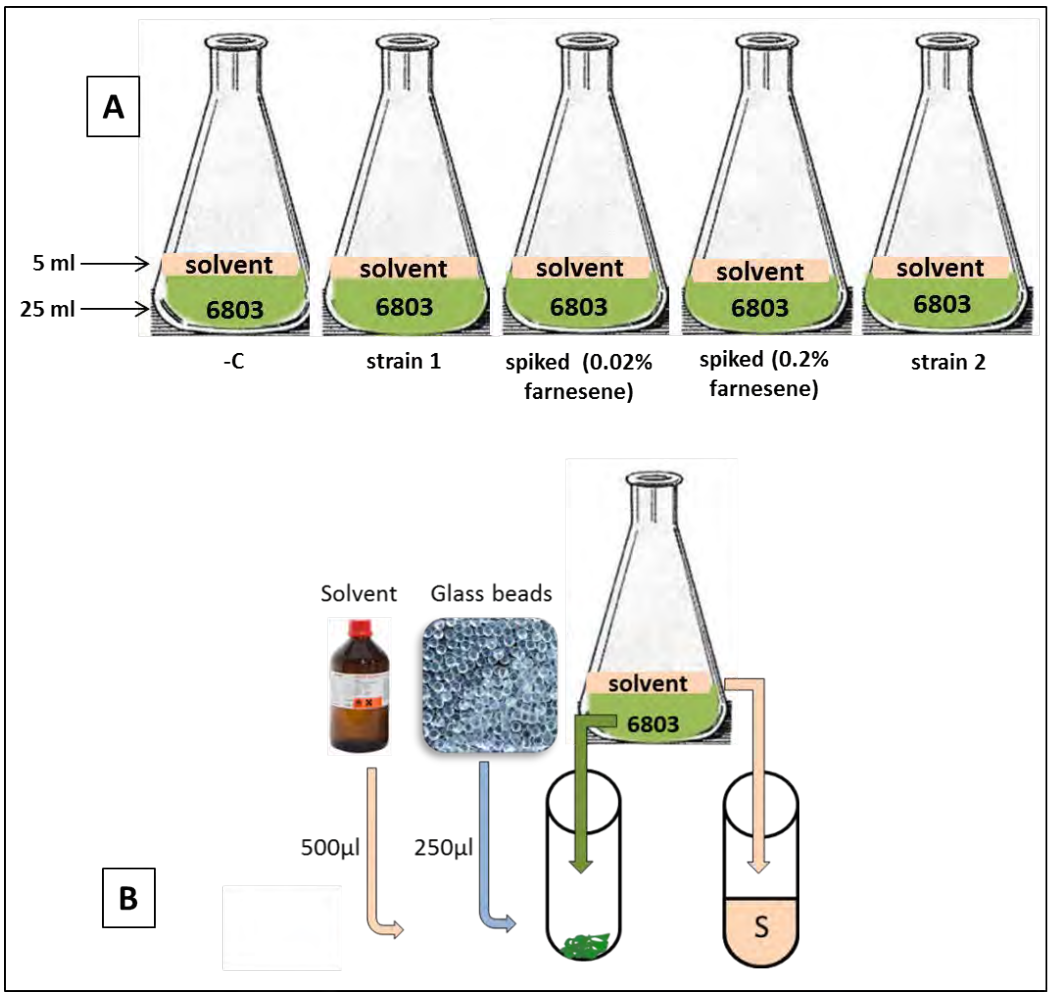


Figure 2.1: Schematic representation of extraction methods

(A) Two-phase extraction method from the cell culture (B) Glass bead extraction of the pellet. N-dodecane and decane were used in both extractions. **Key**: Negative control (-c): WT *Synechocystis* strain transformed with an empty pLAH.A2 vector. Strain 1: *Synechocystis* expressing farnesene synthase (FS) under *psbA2* promoter, Strain 2: *Synechocystis* expressing FS under *nrsB* promoter.

Cells in the remaining culture were pelleted by spinning for 10 min at 16,200 x *g* (Biofuge Pico, Heraeus) and the supernatant discarded. The pellets were re-suspended in 500 µl of solvent (n-dodecane or decane) and transferred to a 2 ml Eppendorf tube containing ~250 µl of acid washed 212-300 µM glass beads (Sigma) Figure 2.1 B. The tubes were vigorously mixed using a vortex for 2 min followed by 2 min incubations on ice. The vortex/ice incubation steps were performed 5 times before spinning the cells for 5 min at 16,200 x *g* and transferring the top solvent phase into GC vials for GC analysis. The samples were run in parallel with standards of 1 ppm, 50 ppm and 100 ppm of commercial β-farnesene synthase prepared in decane and n-dodecane respectively.

2.8.4 Gas Chromatography analysis

For GC analysis for biofuels: Gas Chromatography (GC) analysis on the samples from sections 2.8.1, 2.8.2 and 2.8.3 was carried out using a Clarus 500 GC system equipped with a Clarus 500 mass spectrum (MS) detection system (both from Perkin-Elmer, life sciences, UK). The samples were separated on a Restek Rxi-5ms column, 30 m length, 0.25 mm ID, 1.0 µm film thickness. Helium was used as a carrier gas flowing at a rate of 1 ml/min. 1 µl of each sample was injected using an autosampler (Perkin Elmer), splitless. The oven temperature was initially held at 80^{o}C for 1 min and was increased at a rate of 10^{o}C/min to 250^{o}C where it was held for 1 min. The temperature of the injector and detector were maintained at 260^{o}C and 220^{o}C respectively. A selected ion recording (SIR) mode was used where the following eight masses were selected for: 41, 53, 67, 69, 79, 93, 120, 133. The retention times of product peaks were compared to the trans-β-farnesene standard purchased from Sigma-Aldrich.

For GC analysis of hydrocarbons: quantification of cell extracts from section 2.8.2 was performed by Gas-chromatography using an FID detector (Flame ionization detector). The column used was a Restek 5MS column (30m x 0.25mm x 1μm) with helium carrier gas under a constant pressure of 17Psi. The samples were eluted in an initial temperature of 100 ^{o}C for 1 minute followed by a linear increase of 10 ^{o}C per minute to reach a final temperature after ramping of 280^{o}C which was held for 2 minutes.

2.9 Growth analysis

The growth curve of WT *Synechocystis* was established using an Algem 1.0 photobioreactor (Algenuity, UK) with WT *Synechocystis* grown under a constant temperature of 30^{o}C, a light intensity of 200 µmol/m^{2}/s and constant mixing at 120 rpm. Growth comparisons between different mutants of *Synechocystis* were achieved by growing the strains under the same conditions and monitoring their growth simultaneously using an automated programme. The Algem was set according to the manufacturer's instructions. The OD_{525} was measured and saved automatically every 10 min and a plot constructed using the instrument's built in software when necessary.

2.10 Cell density determination

Cell density measurements were taken by sacrificing the well contents and centrifuging the samples at 4000 *g* for 5 minutes. The supernatant was treated separately and used for GC analysis (6.3.3) while the dry cell weight of the pellets was determined by subtracting the weight of pre-weighted empty 15 ml polypropylene falcon tubes that were used for harvesting the culture from the same tubes containing the pellet after drying the pellets in the oven at 100^{o}C for 24 hours.

Chapter 3: Attempts for the development of a simple system for random insertion mutagenesis in *Synechocystis* 6803

3.0 Introduction

Transformation of the cyanobacterium *Synechocystis* with exogenous DNA (whether homologous or heterologous) is easily achieved with a high level of efficiency (Barten and Lill, 1995). The natural competency of this cyanobacterium is one reason for it being adopted as a model for molecular genetic studies. With the sequencing of its genome (Kaneko et al., 1996) and the development of molecular tools for gene knockouts and gene mutagenesis (Kufryk et al., 2002, Dzelzkalns and Bogorad, 1986, Williams, 1988, Labarre et al., 1989) the identity and function of many gene products have been identified. Nevertheless, out of a total of 3167 potential protein-coding genes in *Synechocystis,* 679 are still with an unknown function (CyanoBase [6]).

Gene modifications represents one approach that truly revolutionised the ways in which protein function is analysed, whether these modifications are introduced at specific loci or randomly in the genome (Vermaas, 1998). Furthermore, the creation of cyanobacterial mutants which are altered in particular functions is one of the most powerful approaches to the study of cellular phenomena such as photosynthesis (Golden, 1988). Proposed techniques for the creation of cyanobacterial mutants using both random mutagenesis and site-directed (targeted) mutagenesis are described in the current and following chapters (chapter 3 and 4 respectively).

Transformation via double homologous recombination, which involves recombination between the chromosome and added heterologous DNA linked to host DNA, is very successful in creating targeted gene knockouts in *Synechocystis* (Kufryk et al., 2002, Vermaas, 1998). This method is illustrated in Figure 4.1 and discussed in detail in chapter 4. It would, however, be advantageous to have a system for introducing heterologous DNA into *Synechocystis* cells directly without the need of creating a plasmid construct in which this DNA is flanked by the chromosomal DNA elements required to mediate homologous recombination. Work presented in this chapter aims to develop a transformation system for the cyanobacterium *Synechocystis* that does

---

[6] http://www. kazusa.or.jp/cyano/

not require the construction of cloning vectors, digestions or ligation steps and which does not rely on homologous recombination to achieve transformation.

According to Kufryk and colleagues, transformation efficiency of the host strain of *Synechocystis* can be increased by two orders of magnitude if a mutant strain is used in which the sll1354 gene encoding the exonuclease RecJ is deleted (Kufryk et al., 2002). In the work presented here, *recJ*– strains were created and employed as a recipient strain for the transformation of a km^R cassette that is 'naked' in the sense that it is not flanked by any DNA homologous to the cyanobacterial genome. The goal of this study is to achieve random insertion of the cassette into the *Synechocystis* genome, and thus develop this as a method for random mutagenesis in which mutated genes are 'tagged' by the km^R cassette. Thus, unlike conventional UV or chemical mutagenesis, the mutated gene can be rapidly identified using inverse PCR that amplified genomic DNA flanking the cassette, or cloning approaches that rescue the cassette plus flanking sequence back to *E. coli.* Figure 3.1 shows an illustration of the proposed random insertion mutagenesis method.

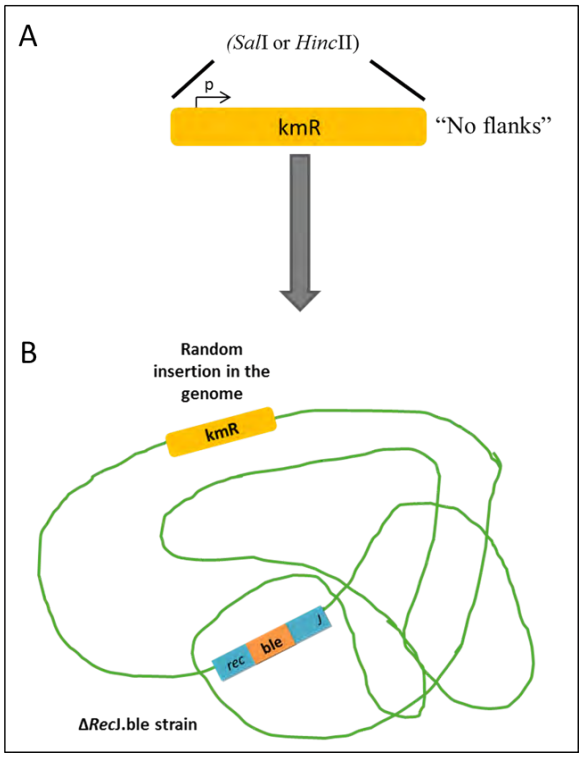


Figure 3.1: Illustration of the proposed method of random km^R cassette integration into a *recJ*- strain.

A) Isolation of naked km^R cassette from pUC4K plasmid as *Sal*I and *Hinc*II fragments with no flanks. (B) Transformation of the naked km^R cassette into Δ*rec*J.ble strain achieving random integration within the genome. **Key:** ble; bleomycin selectable marker.

3.1 Background

3.1.1 Current random methods of mutagenesis

In contrast to targeted gene disruptions (so-called 'reverse genetics') in *Synechocystis* which are based on double homologous recombination, random methods of disruption of the host DNA includes chemical and UV mutagenesis and transposon mutagenesis (also referred to as classical mutagenesis or 'forward genetics') (Golden, 1988). Both chemical and UV mutagenesis have been of limited value in *Synechocystis* possibly due to the availability of multiple genome copies which makes the detection of recessive mutations induced by these methods more complicated (Labarre et al., 1989).

In heterotrophic bacteria, using transposons for mutagenesis is the most common approach for identifying genes that have not been described before. However, the use of transposons for mutagenesis in cyanobacteria is not well characterised (Labarre et al., 1989, Golden, 1988). Transposon-based methods that have been developed for use in other strains can work poorly in cyanobacterial hosts (Berla et al., 2013, Golden, 1988). Although the reasons behind this have not been fully elucidated, it is reported that the method results in only few mutants (Tandeau de Marsac et al., 1982). Hence, the ability to find conditions that will induce transposition at a higher frequency or the identification of elements that readily transpose in cyanobacteria strains would enhance the usefulness of this method (Golden, 1988). As a result, other methods were recently developed that do not depend on classical mutagenesis methods. One example is the insertional mutagenesis method developed by (Labarre et al., 1989). In this method, the disruption of the host genome is done by the insertion of an antibiotic cassette. The advantage here is that recessive mutants can be recovered since positive selection for the antibiotic allows recovery of homoplasmic mutants in which all copies of the genome carry the recessive mutation. Random integration of the marker DNA was achieved by randomly ligating it to a pool of restriction fragments of the host genome. The ligation mix (selectable marker physically linked to cyanobacterial DNA) is then used directly for transformation into *Synechocystis*. The authors highlight that this does not suggest that the integration of the antibiotic marker necessarily results from homologous DNA pairing and recombination between the donor DNA molecule and the host

genome. Rather, the linear marker DNA might need to be extended by adding additional DNA to either end merely to protect against exonucleolytic degradation, whereas integration itself may need little or no homology to the host DNA. It has been suggested that short exposure of *Synechocystis* to low-level UV irradiation reduces the effect of endogenous nucleases and allows for integration of heterologous DNA into the cyanobacterial host independent of either autonomously replication or homologous recombination (Dzelzkalns and Bogorad, 1986). Nevertheless, the process still involves the use of UV irradiation, which as mentioned earlier, has a number of disadvantages.

From the previous discussion, it appears reasonable to speculate that if the native host nuclease degradation process can be mitigated, it might be possible to transform the host cell with heterologous DNA without the need for the marker DNA to be physically linked to *Synechocystis* chromosomal DNA. In *Synechocystis*, this could possibly be done by deleting the *recJ* gene. The importance of eliminating the exonuclease RecJ for better transformation efficiency was identified by Porter (Porter, 1986) who also suggested that the most pressing need in cyanobacterial genetic engineering in general is the development of *rec*- strains of the appropriate organisms.

The presence of the exonuclease RecJ in *Synechocystis* is well documented. This 759 amino acid protein is a single-strand DNA-specific exonuclease that degrades foreign DNA from the 5'-3' end. It was Kufryk and colleagues who were the first to develop a *recJ*- strain in *Synechocystis* 6803 (Kufryk et al., 2002). In their work, it was demonstrated that deleting the sll1354 gene that encodes RecJ significantly improved the transformability of the host strain of *Synechocystis* by two orders of magnitude. It would appear that loss of the exonuclease extends the half-life of the transforming DNA allowing inefficient illegitimate (*i.e.* non-homologous) recombination events to occur.

3.2 Chapter aims

- To develop a rapid and simple method for random foreign gene insertion into the genome of *Synechocystis* based on an antibiotic resistance marker and a RecJ- deficient host as illustrated in Figure 3.1.
- To use the newly established method to identify novel mutants for identification of novel mutants using forward-genetic screens and also as a

promoter-trap when employing a promoter-less version of the marker to identify novel strong or inducible/repressible promoters.

3.3 Results and discussion

3.3.1 Creation of the Δ*recJ Synechocystis* mutant

For the creation of the *Synechocystis* mutant strain with a deleted *recJ* gene, an existing pRecJ::ble plasmid construct (kindly provided by Saul Purton) was utilised. In this construct *recJ* is interrupted with the *ble* marker gene conferring resistance to zeocin. Details of the plasmid can be found in Appendix 1. The transformation method for wild type *Synechocystis* with the pRecJ::ble plasmid is described in section 2.4.2.

Putative colonies appeared within ten days post antibiotic selection and eight isolated colonies were taken through three rounds of re-streaking on selective medium to ensure that the transformants reached the homoplasmic stage. Genomic DNA from eight lines was isolated as described in section 2.3.3.1 and used in a PCR reaction with primers annealing at the 5' and 3' ends of *recJ* (Appendix 3) to confirm the insertion of the *ble* gene. Various controls were also included: namely the pRecJ::ble plasmid which was used a positive control, the plasmid pRecJ that contains an intact version of *recJ*, and genomic DNA from wild-type *Synechocystis* 6803 which both served as negative controls. This was done to compare the band sizes of the wild type vs. disrupted versions of *recJ* to confirm the isolation of successful transformants. The expected size for the WT PCR product is 2002 bp and that of the *recJ*::ble allele is 1695 bp as illustrated in Figure 3.2.A.

As shown in Figure 3.2.B, the results of the PCR confirm that transformant lines T1 – T4 contain a disrupted version of *recJ*. Furthermore, the lack of a 2 kb wild type (WT) band indicates that all 12 or so copies of the polyploid genome present in each cell have the mutant allele (*i.e*. the transformant lines are homoplasmic for the knockout). Since all the transformant lines are presumed to be genetically identical, only one was chosen as a representative line and its phenotype confirmed by growth tests on different solid media as shown in Figure 3.3. Such 'spot tests' were undertaken as described in 2.2.6 and confirmed that the line is resistant to zeocin while remaining sensitive to kanamycin and can thus be used as a recipient strain when using the km^R marker.

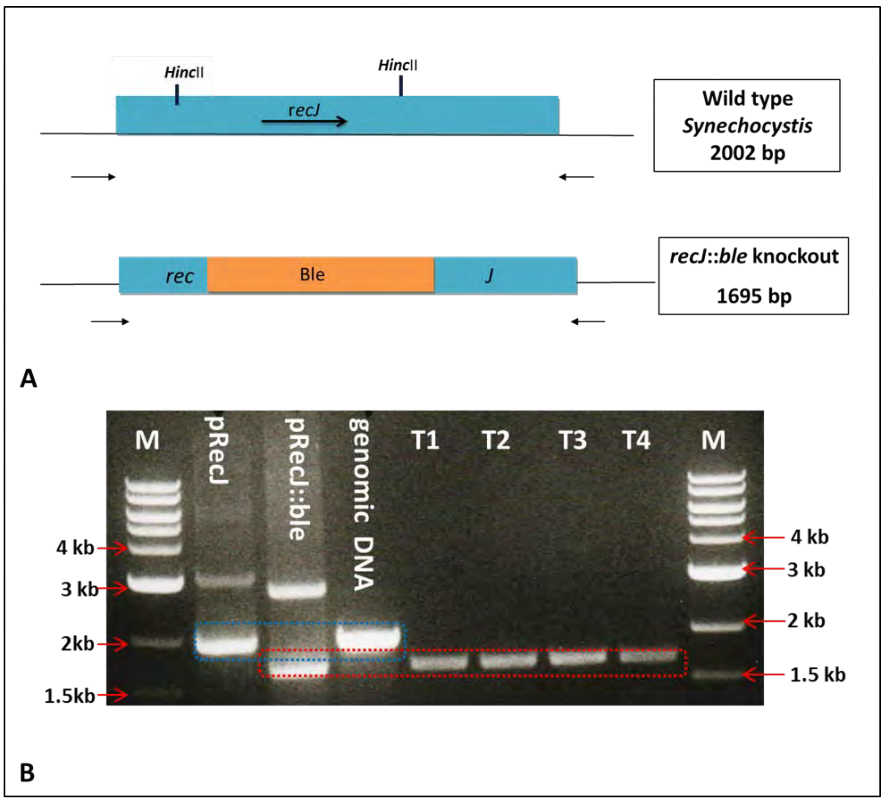


Figure 3.2: PCR confirmation for the isolation of a Δ*recJ Synechocystis* mutant

(A) Schematic representation of WT size of the *recJ* gene (2.0 kb) versus the *recJ*::*ble* knockout (1.7 kb) (B) PCR confirming the isolation of Δ*recJ Synechocystis* transformants (T1 – T4). **Key:** pRecJ; plasmid containing uninterrupted version of the recJ gene (-C), pRecJ:ble; plasmid containing recJ gene deleted with a ble marker (+C), T1-T4; transformants.

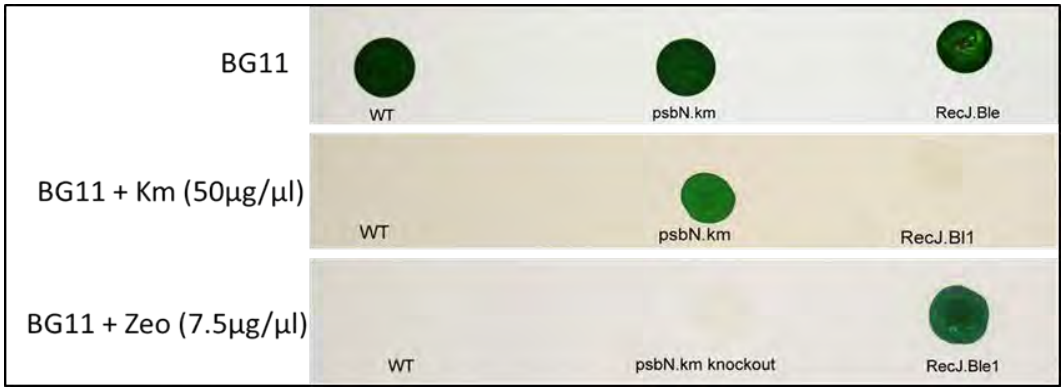


Figure 3.3: Spot tests of WT, a kanamycin-resistant transformant line (*psbN*.km) and a representative *recJ::ble* transformant line.

3.3.2 Transformation of the Δ*recJ Synechocystis* mutant

To test for random integration of foreign DNA into the Δ*recJ* mutant strain, a kanamycin resistance cassette was isolated by digesting the pUC4K plasmid (Appendix 1) with *Sal*I according to methods described in section 2.3.11.2. The

1.2 kb band corresponding to km^R cassette was excised from the gel and purified. The same method was performed to isolate the km^R cassette as a *Hinc*II fragment.

The cassettes, isolated with blunt ends (*Hinc*II) or a four-base 5' overhang (*Sal*I), were both used for two independent transformation experiments using either the wild type strain or the Δ*recJ* strain. After ten days, many colonies were visible for the Δ*recJ* strain while only very few small colonies were obtained in the wild type recipient strain for both isolated km^R cassettes as shown in Figure 3.4. The tiny colonies that appeared from the transformation of WT strain were too small to be picked and appear to be pseudo-transformant colonies arising through depletion of antibiotic in the media. These colonies were not analysed further. Similar early work (Labarre et al., 1989) showed that the transformation of selectable markers that were not physically linked to *Synechocystis* DNA (*i.e*. carrying no homology to the genome of the cyanobacterial host) produced no transformants in wild type *Synechocystis* strains.

Colonies were readily obtained using the *recJ* mutant strain indicating that in this mutant background the marker DNA is sufficiently stable to become incorporated into the genome, and that DNA integration does not require tracts of homology between the DNA and the genome. Sixteen isolated colonies were taken through three rounds of re-streaking on selective media to ensure they reached the homoplasmic state before being tested by PCR analysis. Representative transformants for both the *Hinc*II and *Sal*I forms of the cassette were included to test whether using blunt or sticky ends would have an effect on the integration pattern of the selectable marker within the genome.

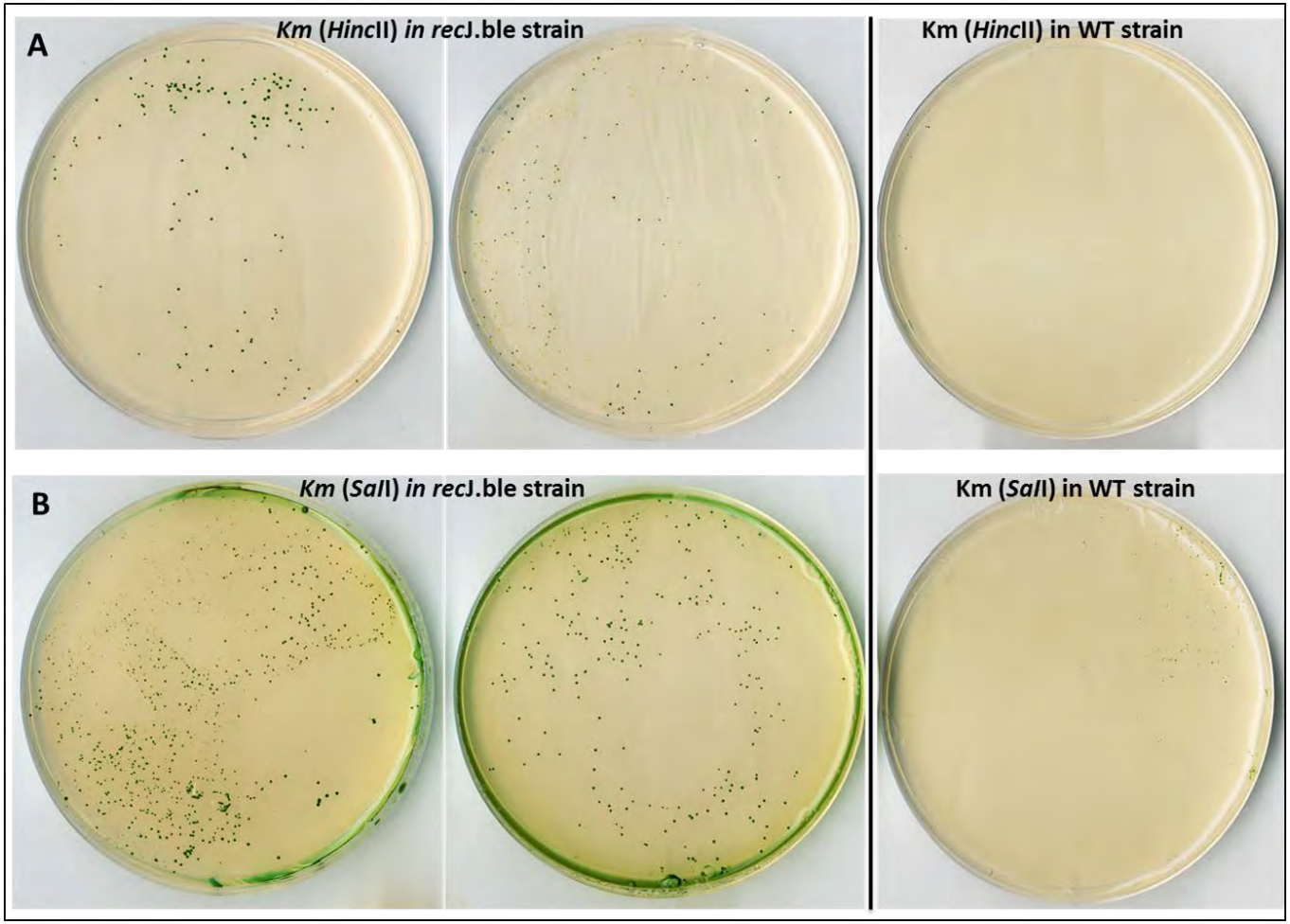


Figure 3.4: Transformation of WT and Δ*recJ* strains with the kanamycin-resistance cassette

Colonies obtained from the transformation of Δ*recJ* and WT *Synechocystis* strains with the km^R cassette as a *Hinc*II fragment (A) and as a *Sal*I fragment (B).

3.3.3 Analysis of Δ*recJ Synechocystis* transformed with the km^R cassettes

3.3.3.1 PCR analysis

DNA preparations of the isolated transformants obtained in section 3.3.2 were used in a PCR reaction to test for the integration of the km^R cassettes into the transformants genome. Primers used for the amplification of the km^R cassette in this PCR reaction are listed in Appendix 3. A positive control for the cassette (plasmid pUC.psbN.km^R) and negative controls with DNA lacking the cassette (genomic DNA from the WT *Synechocystis* and the *recJ::ble* recipient line) were also included in the PCR reactions to test the specificity of the primers. The results are presented in Figure 3.5 and show a band of 1.2 kb band in all eight of the '*Hinc*II' transformant lines confirming the presence of the km^R cassette. The same results were obtained when analyzing the eight transformants generated with the km^R cassette as a *Sal*I.

It is worth noting here that it is not advisable to use Phusion® High-Fidelity DNA polymerase, for the screening of transformants for the presence of km^R cassette. It was observed that a band corresponding to the km^R cassette was routinely detected in the negative control samples. After prolonged investigations to identify the source of contamination, it appeared that traces of the km^R marker were present in the polymerase itself. This is probably due to the fact that the recombinant Phusion is expressed in *E. coli* from a plasmid carrying the kanamycin-resistance gene and the isolated polymerase is contaminated with trace amounts of such plasmid. For this reason Taq polymerase was used for screening tests instead.

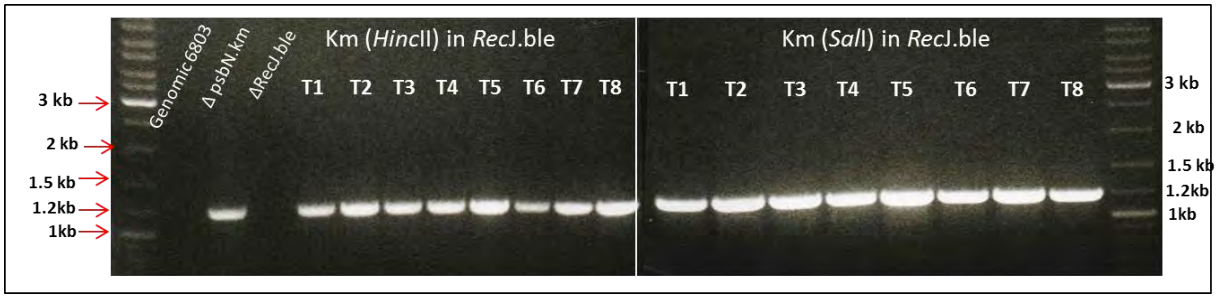


Figure 3.5: PCR confirmation of km^R integration in the Δ*recJ* strain.

All transformants show an amplified band for the km^R cassette at ~1.2 kb indicating the presence of km^R cassette as both *Hinc*II and *Sal*I fragments in all isolated transformants. **Key:** ΔpsbN.km: plasid containing a *psbN* gene interrupted with a km^R gene (+C), ΔRecJ.ble: plasid containing a *recJ* gene interrupted with a ble selectable marker (-C).

3.3.3.2 Southern blot analysis

Whilst the PCR result confirms the presence of the km^R cassette in the transformant genomes, it does not provide any information about the copy number of the cassette or the site of integration in the genome. In order to determine whether the cassette integrates at random loci, as predicted, genomic DNA was prepared for the 16 isolated transformants as detailed in section 2.3.3.2 in order to carry out a Southern blot. In order to simplify the expected pattern and to test for random integration, two restriction enzymes (*Xmn*I and *BsaA*I) that do not cut the km^R cassette were chosen for digestion. Accordingly, the size of any genomic fragment containing the cassette will be dependent on the particular insertion site and the location of *Xmn*I or *BsaA*I sites around the site. If a single copy of the cassette has integrated randomly in each transformant then we would predict a Southern blot result in which a single band hybridizing to the probe is seen for each transformant line, but with each band of a different size.

The Southern blots were set up and the probe was prepared according to methods detailed in section 2.5.1 and 2.5.2, respectively. Figure 3.6 shows the results with DNA isolated and digested from 15 transformants (one sample was lost during genomic DNA preparation) and probed with a labelled Km^R cassette isolated from pUC4K. Both the *Xmn*I and *BsaA*I digestions show the same pattern for all transformants – namely, a band of the same size in each transformant, suggesting that the marker is either integrating a single copy into a preferred location within the genome, or that the DNA has not integrated but is being freely replicated, either as a linear molecule or in a circular form. This observation was contradictory to what was expected. Therefore further investigations were made in order to determine if the DNA is freely replicating within the cell or integrated into the genome, and if it was the latter, the site of integration.

To investigate whether the marker is present as a freely replicating form, a further Southern blot was set up where enzymes *Xmn*I and *BsaA*I were used to digest the genomic DNA of two independent transformants Δ*recJ*.ble.km(*Hinc*II) and Δ*recJ*.ble.km(*Sal*I), while one genomic sample was left undigested (control). A Δ*psbN*.km transformant in which the *psbN* gene was interrupted by the insertion of the km^R cassette (Chapter 4) was used as a positive control to check the functionality of the probe. Figure 3.7.A shows the possible scenarios of the

digested genomic products containing the km^R cassette. Since the cassette lacks both restriction sites then 'digestion' of a freely replicating form (either linear or circular) should have no effect and the distance migrated on the blot will be the same for all three cases (uncut, *Xmn*I-cut and *Bsa*AI-cut). However if the genomic DNA digested with *Xmn*I or *BsaAI* in the two independent transformants resulted in different band sizes to each other and to the uncut control, this would indicate that the cassette had integrated into the host genome – either into the main chromosome or into one of the host plasmids.

As seen in Figure 3.7.B two different band sizes are seen for the digestion of *Xmn*I and *BsaA*I confirming that the km^R cassette was integrated into the genome of the recipient strain. Importantly, the fact that the same size bands are seen for independent transformant lines following digestion with a particular restriction enzyme (Figures 3.6 and 3.7) suggests that integration is occurring at a specific genome locus rather than at random sites. The next stage was therefore to identify the exact location at which the cassette was integrating in the host genome.

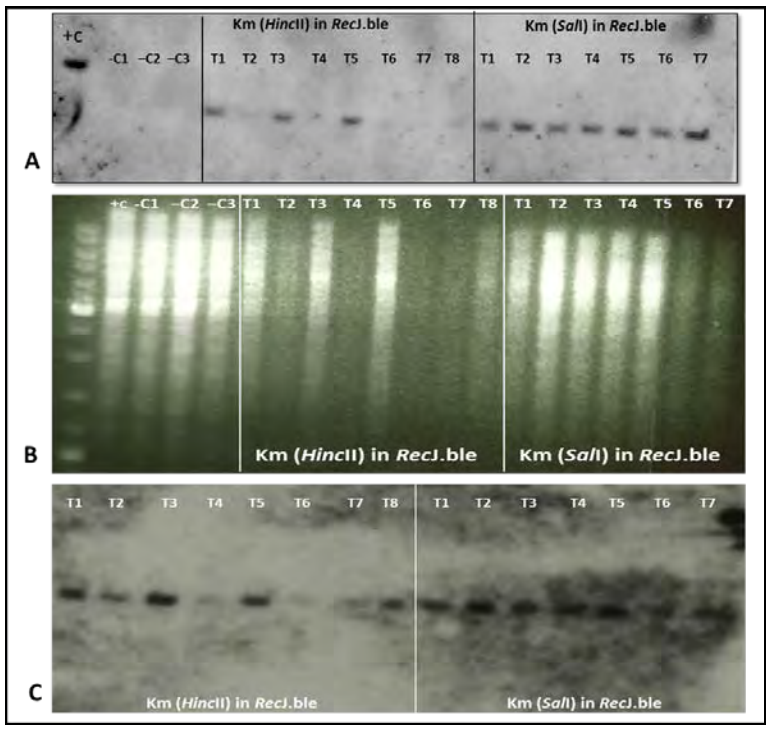


Figure 3.6: Southern blot analysis of Δ*Rec*J.ble transformants.

Southern blot analysis shows presence of the km^R cassette in the genome of all transformants. (A) Blot from DNA digested with *Xmn*I. (B) Stained gel of *Xmn*I digests. Note: the absence of some bands in the Southern blot can be explained by the low yield of the genomic DNA extracts. (C) Southern blot analysis of transformants following digestion with *Bsa*AI. **Key:** Δ*psbN*.km transformant (+C), genomic DNA of recJ.ble strain (-C1, -C2), genomic DNA of WT6803 strain (-C3). T1-T8 transformants containing the km^R cassett.

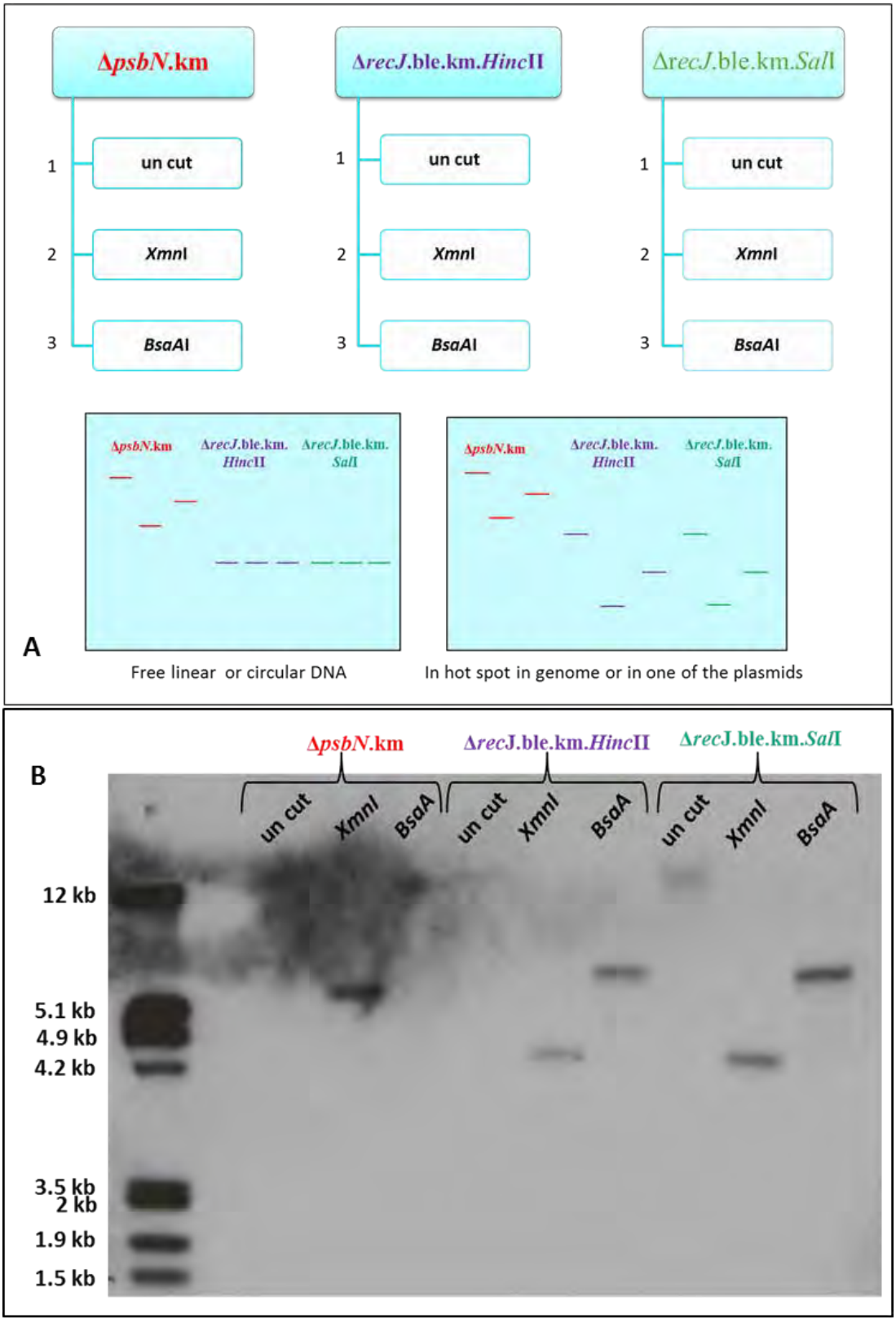


Figure 3.7: Southern blot to test for genomic integration of km^R in the Δ*recJ* strain.

(A) Schematic representation of the possible ways the naked km^R cassettes are being inserted into the Δ*recJ* recipient strains. (B) Southern blot analysis confirms that the cassette has integrated into the genome of the recipient strain.

3.3.3.3 Identifying the site of integration

In order to identify the exact location at which the km^R cassette had integrated, it was necessary to isolate the cassette together with flanking genomic DNA from the transformants, and clone it into a plasmid vector (*e.g.* pJET). This would allow the sequencing of the flanking DNA and identification of the insertion site. For this, firstly a set of eight enzymes (*Age*I, *Bam*HI, *Bsp*EI, *Bss*HII, *Eae*I, *Sph*I, *Xba*I and *Eco*RV) that do not cut the km^R cassette were selected to be used for the digestion of the genomic preparations of a representative Δ*recJ*.ble.km(*Sal*I) transformant. This was done to see which of the enzymes would result in the smallest band size containing the km^R cassette that could be used for cloning purposes into pJET. The eight samples of digested genomic DNA were then used in a Southern blot that was probed using the km^R probe. Results from the blot showed that the enzymes *Xba*I and *EcoR*V yielded the smallest bands containing the km^R cassette (~1.4 and ~4.5 kb, respectively) as seen in Figure 3.8. This information was used to prepare larger scale genomic digests of the same transformant using these two enzymes as per section 2.3.11.2.

The digests were run on a 1% agarose gel and the area around each band size was excised and purified as shown in Figure 3.9.A. The purified gel extracts were then cloned into pJET and transformed into competent *E. coli* cells (section 2.4.1) that were selected on LB plates supplemented with both ampicillin at 75 µg/ml (to select for uptake of the pJET DNA) and kanamycin at 50 µg/ml (to select for the Km^R cassette ligated into pJET). No colonies were obtained for the *Eco*RV cloning, but eight colonies were recovered for the *Xba*I cloning (this difference is possibly due to the fact that cloning of smaller fragments into plasmids is more efficient).

To confirm the presence of the km^R cassette in the plasmids isolated from the eight colonies, test digests were done using *Xba*I. This was because *Xba*I cuts pJET once and does not cut the km^R cassette (the sites were lost during cloning into the vector due to the use of a blunt cutting enzyme as part of the protocol), therefore a successful cloning would show a single band of ~4.4 kb comprising the 3.0 kb vector plus the 1.4 kb insert.

Figure 3.9.B shows that a 4.4 kb band is obtained for all six of the plasmids analysed. Plasmids 4 and 6 were sent for sequencing using pJET-specific

primers flanking the cloning site in order to determine the genomic sequence at either end of the cloned km^R cassette and thus identify the site of integration.

Figure 3.8: Southern blot of genomic DNA digested with eight enzymes and probed with a km^R probe.

Results showed that the enzymes *Xba*I and *EcoR*V yielded the smallest bands containing the km^R cassette (~1.4 and ~4.5 kb, respectively).

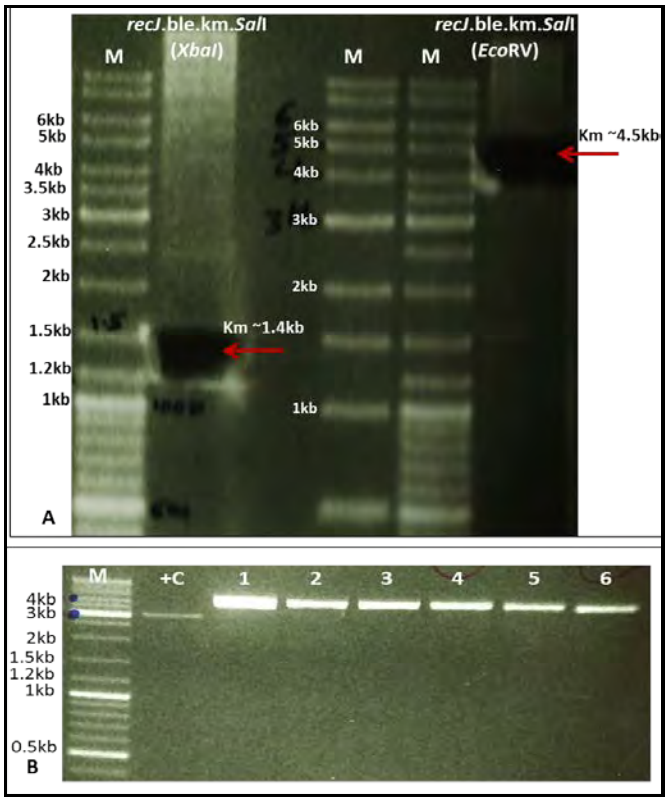


Figure 3.9: km^R cassette extracted as *Xba*I and *EcoR*V fragments and cloned in p.JET.

(A) Gel extraction of the km^R cassette isolated by genomic digestion with *Xba*I and *Eco*RV enzymes. (B) Test digests using *Xba*I confirm the presence of the km^R cassette in the six plasmids analysed. +C represents the empty pJET vector (3.0 kb).

3.3.3.4 Sequence analysis

Analysis of the DNA sequence from the Δ*recJ*.ble.km(*Sal*I)#4 transformant confirmed that the plasmid insert contained the km^R cassette together with flanking regions of 152 bp and 47 bp. These two sequences were used in a BLASTN search of the published *Synechocystis* genome sequence [7]. The 152 bp sequence showed a match to part of the *speA* gene (gene code = slr1312) that encodes for an arginine decarboxylase (Figure 3.10.A and B). However, the downstream element (47 bp) showed no significant match to the genome of *Synechocystis* possibly due to the short length of the query sequence (Figure 3.10.C). Nonetheless, the genome around the *speA* gene was examined and predictions were made as to whether the km^R cassette has been inserted in between *speA* and the upstream gene *psbAII* or had in fact replaced part of the *psbAII* gene as depicted in Figure 3.10.C. The sequence analysis of the cassette region from the other transformant (Δ*recJ*.ble.km(*Sal*I)#6) confirmed the same integration pattern for the km^R cassette – *i.e.* upstream of the *speA* gene, but again no additional information was obtained on the location of the other end of the cassette.

In order to further study the integration of the cassette in the different transformant lines primers were designed (Appendix 3) in order to amplify the area of the *Synechocystis* genome containing the *speA* and *psbA*II genes. The location of the primers (named P1 and P2 for simplicity) is indicated on Figure 3.10.C. The expected wild type band size using these primers would be ~1.5 kb while transformants containing the km^R cassette at the locus were expected to show a band size of ~2.7 kb if the 1.2 kb cassette has inserted without replacing part of the locus.

All 16 transformants together with an untransformed control strain were analysed by PCR with the results shown in Figure 3.11. The results show a band shift from the wild-type in all transformants tested. However, the band shift reveals a size increase of only ~100 bp as opposed to the expected 1.2 kb. This suggests that the km^R inserted in this area was actually replacing approximately 1.1 kb of the genome including part of *psbAII* as shown in Figure 3.10.C. The *psbAII* gene,

---

[7] http://genome.microbedb.jp/cyanobase/Synechocystis

along with two other genes *psbAI* and *psbAIII,* encode the highly expressed D1 protein of photosystem II. The *psbAII* gene is dispensable and its deletion is not lethal to the cell (as seen in these experiments). What makes the *psbAII* gene dispensable is the fact that the *psbAIII* gene express enough D1 protein in the absence of *psbAII* (Mohamed et al., 1993).

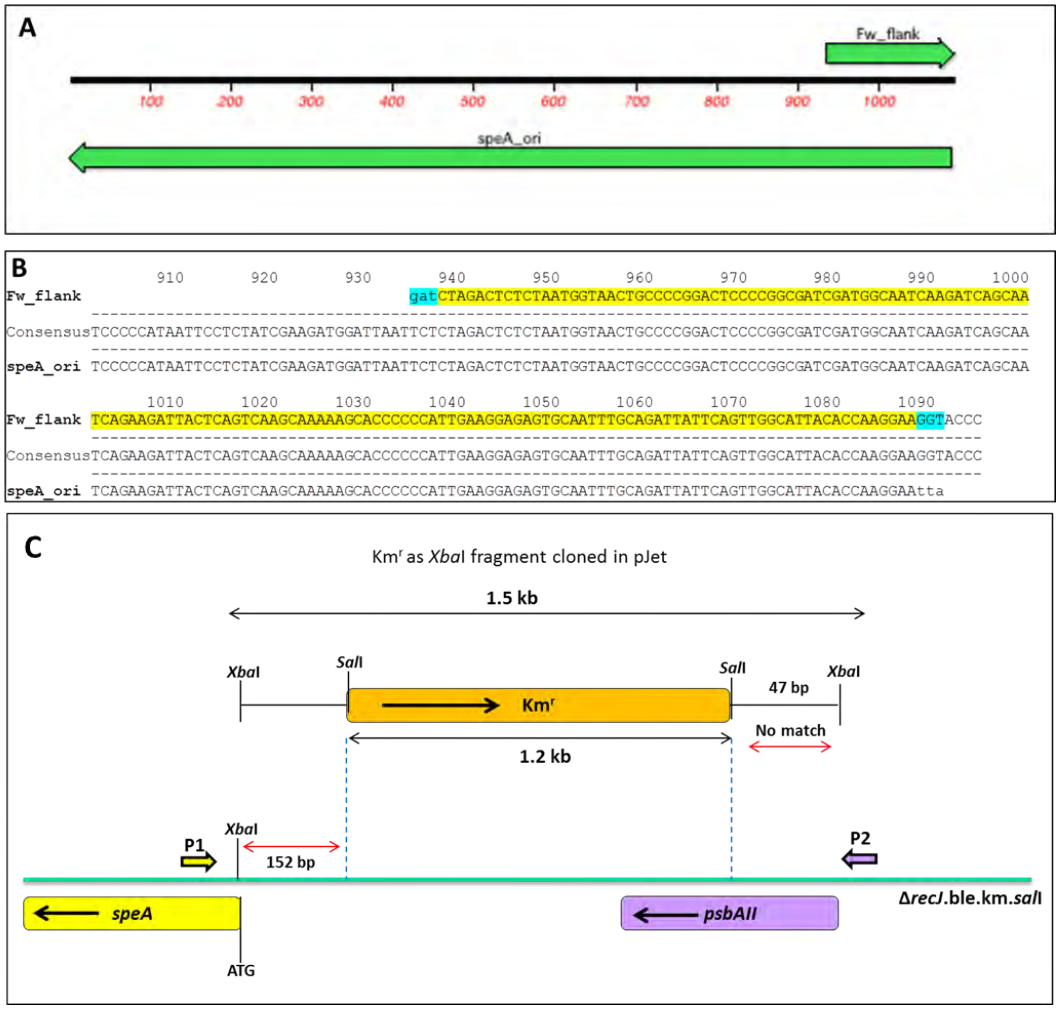


Figure 3.10: Sequance alignment and schematic representation of the Δ*recJ*.ble.km(SalI) transformant.

(A) Figure showing the alignment between the forward flank sequence (152 bp) and part of the *speA* gene and their orientations. (B) Sequence alignment between the 152 bp sequence (highlighted in yellow) and part of the *speA* gene. (C) Schematic representation of the Δ*recJ*.ble.km(SalI) transformant. P1 and P2 represent primers used for amplification of target area (*speA* and *psbA*II genes).

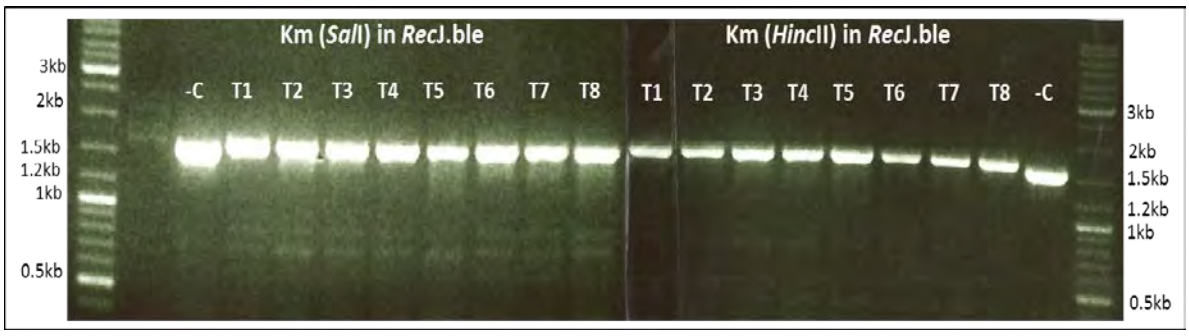


Figure 3.11: PCR results confirming the presence of km^R cassette in the *speA*/*psbA*II locus.

PCR results show an increase of ~100 bp of DNA into the *speA.psbAII* locus in all the 16 transformants. -C: WT genome (1.5 kb), transformants (~1.6).

In order to confirm that the km^R cassette was replacing part of the genome at the *psbAII* locus, a new PCR analysis was carried out using a primer designed further upstream of *psbAII* within the adjacent gene sll1212, and another primer within the cassette itself, named P3 and P4 for simplicity, as shown on Figure 3.12 (primers listed in Appendix 3). This was also done to try to gain a better understanding as to why the km^R cassette had not inserted at random sites in the genome of the transformants. One transformant was selected for analysis and the PCR product obtained was purified and sent for sequencing using primers listed in Appendix 3.

Sequence results revealed that a recombination event took place in which the km^R cassette replaced the majority of the *psbAII* gene. Figure 3.12 shows a physical map of the DNA surrounding the insertion site of the km^R marker gene in both the wild type (Δ*recJ*.ble strain in this case) and the Δ*recJ*.ble.km mutant. The observation that genomic DNA had been deleted in the transformants is similar to earlier findings by two groups (Chauvat et al., 1989, Labarre et al., 1989) who showed that transformants obtained using km^R and cm^R cassettes that had been randomly ligated to fragmented *Synechocystis* DNA, were usually accompanied by a gene deletion. The annotated sequence of the Δ*rec*J.ble.km.*Hinc*II transformant can be found in Figure 3.13.

It is unclear whether there is related sequence to the *speA*/*psbA2* sequence within the km^R cassette that favours its integration in this locus, thus preventing its random integration into other sites in the genome, although a BLASTn search of the genome using the cassette sequence failed to identify any homology. On the other hand, it may be possible that there is something related to the

sequence of that area of the genome, such as the presence of a strong promoter or a specific DNA element that makes it act as a “hot spot” for integration. It has been reported by (Zang et al., 2007) that a mechanism exists in *Synechocystis*, based on a palindromic element of the genome, that allows for site specific recombination to take place, a property that leads to DNA integration taking place easily at some “ hot spots” of the genome. Two of these elements, namely, the HIPI (highly iterated palindrome) and IS (insertion sequences) have already been identified (Ikeuchi and Tabata, 2001). There are 3160 copies of the 8 bp HIPI palindromic sequences, GCGATCGC that is found in a frequency of once every 1131 bp while there are 77 IS-like elements (grouped into nine groups based on sequence similarities) also present throughout the genome. These elements may play a role in the local dynamics alteration of the genome or mediate homologous recombination between IS-like sequences at different positions in the genome. Examining the genome around the integration site shows the presence of the 8 bp HIPI palindromic sequences both upstream and downstream the *psbAII* site (Appendix 6). This may be the reasoning behind the integration of the km^R cassette at this location.

However, in order to test the two possibilities, a further experiment was designed in which a recipient line is created where the *recJ* gene was interrupted with the km^R cassette (rather than with the *ble* cassette), and then a ‘naked’ *ble* cassette (with no flanking sequence) is used as the foreign DNA to transform the new host. The integration pattern of the *ble* marker gene in the new strain would provide more information in order to come to more conclusive results. Such work is presented in the following sections.

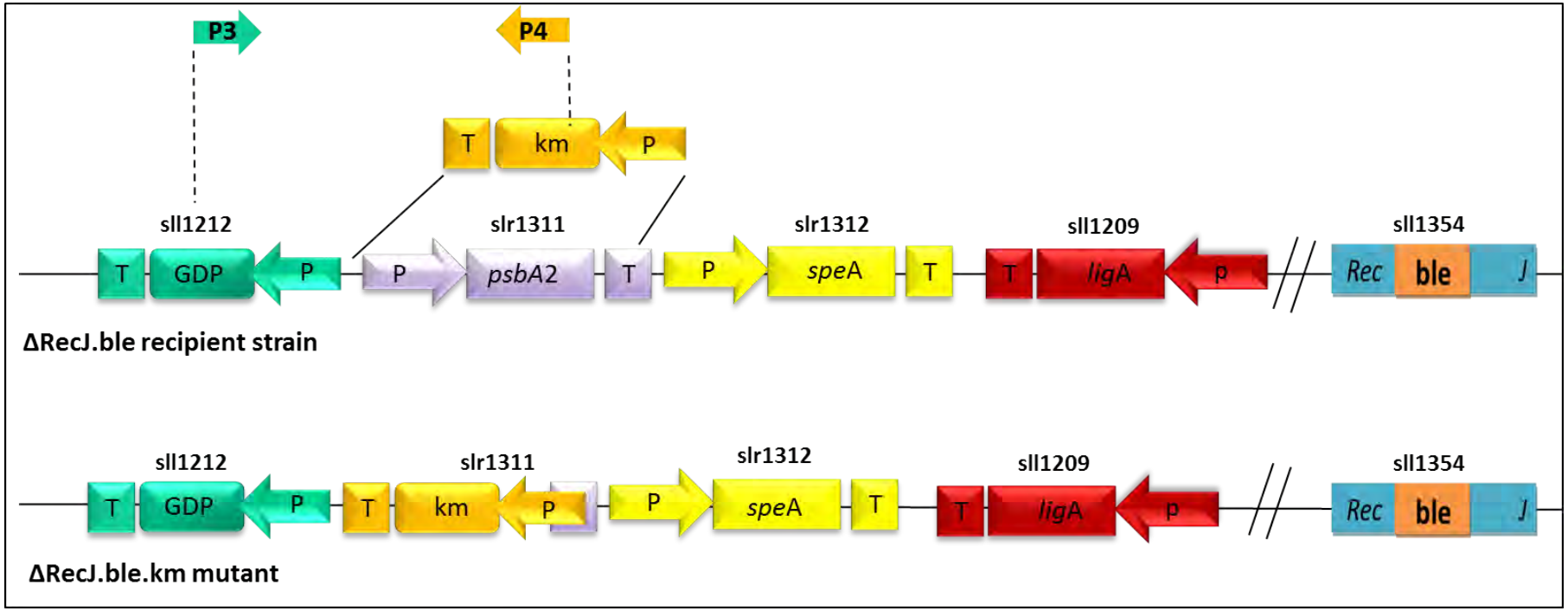
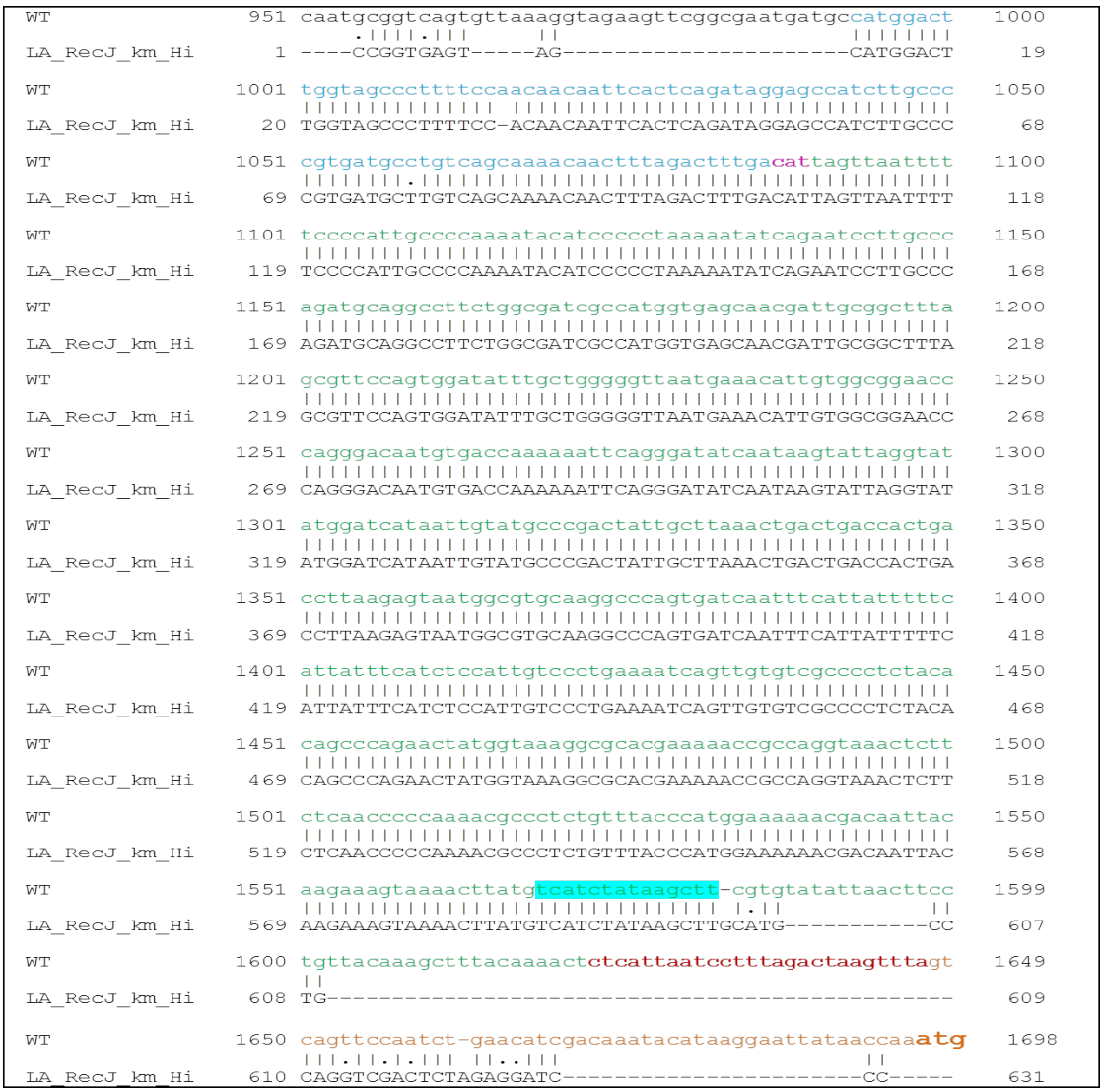


Figure 3.12: Illustration of the insertion of the km^R marker in the *psbAII* locus.

Comparison of the map of the DNA surrounding the insertion site of the km^R marker gene in wild type (top) and in the Δ*rec*J.ble.km mutant (bottom). The km^R marker replaced most the *psbA*II locus. Gene codes are given above each gene. P3 and P4 stand for primers 3 and 4 used for PCR analysis.

A



KEY: GDP-mannose (sll1212) (blue), GDP-mannose start codon (pink), GDP-mannose/*psbAII* intergenic sequence (green), *psbAII* promoter (maroon), part of *psbAII* CDS including start codon (orange). Blue highlighted sequence is the site after which the km^R cassette was integrated.

B

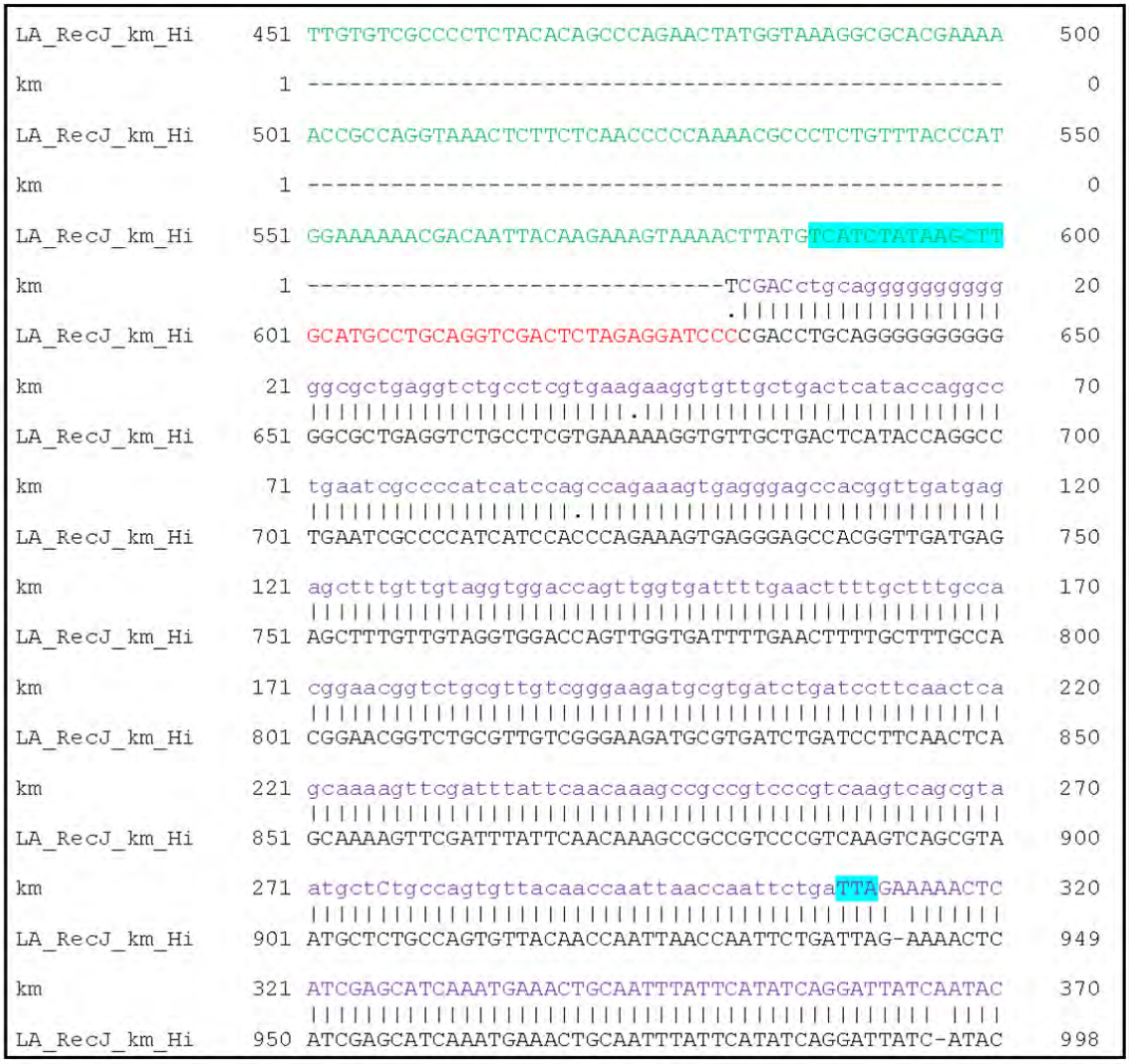
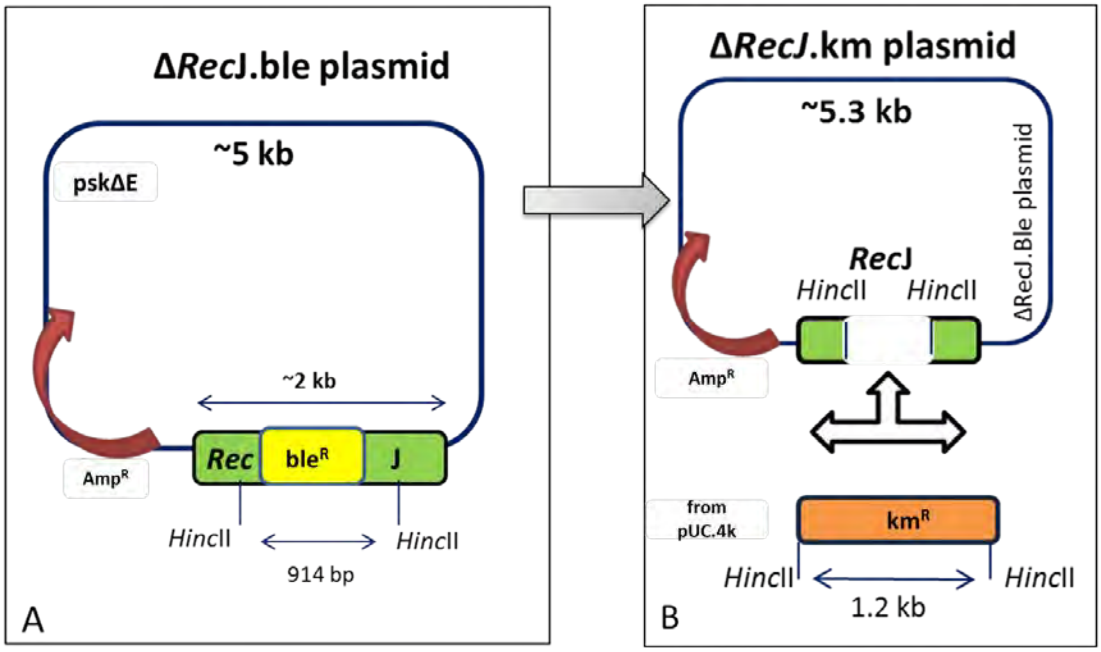


Figure 3.13: The annotated sequence of the Δ*rec*J.ble.km.*Hinc*II transformant.

(A) Sequence alignment between WT (Δ*rec*J.ble) and Δ*rec*J.ble.km.*Hinc*II transformant show the absence of the *psbAII* sequence in the mutant strain. **(B)** Sequence alignment between the km^R cassette and the Δ*rec*J.ble.km.*Hinc*II transformant shows integration of the km^R cassette just upstream of the *psbAII* locus. The recombination resulted in the replacement of the *psbAII* sequence and a DNA rearrangement indicated by the presence of 31 bp of unidentified DNA at the site of integration (red). **KEY**: GDP-mannose (sll1212) (blue), GDP-mannose start codon (pink), GDP-mannose/*psbAII* intergenic sequence (green), *psbAII* promoter (maroon), part of *psbAII* CDS including start codon (orange), recombination site (red), km^R gene (purple). Blue highlighted sequence is the site after which the km^R cassette was integrated.

3.3.4 Creation of the Δ*rec*J.km *Synechocystis* mutant.

For the creation of a new *recJ*- strain of *Synechocystis,* the pRecJ::ble plasmid was modified such that the *recJ* gene was disrupted with the km^R cassette (as opposed the *ble* cassette). The cloning strategy for the construction of the pRecJ::Km plasmid is shown in Figure 3.14.



Figuer 3.14: The construction of the pRecJ::Km plasmid.

(A) The ΔRecJ.ble plasmid was digested with *Hinc*II to drop out 914 bp of the *recJ* gene containing the *ble* cassette. (B) The purified digested plasmid was ligated to a km^R cassette isolated from pUC4K as a *Hinc*II fragment (1.2 kb) resulting in the construction of the ΔRecJ.km plasmid (5.3 kb).

The pRecJ::Km plasmid was used for transformation wild type *Synechocystis* as described in section 2.4.2. The transformants were selected by supplementing the BG11 plates with kanamycin at a concentration of 200 µg/ml. Putative colonies appeared within five days post antibiotic selection and four isolated colonies were taken through three rounds of re-streaking on selective medium to ensure that the transformants were homoplasmic. Genomic DNA was isolated from the putative transformants as described in section 2.3.3.1 and PCR reactions were performed as described in section 2.3.5 using primers to the *recJ* gene as described in section 3.3.1. Controls using DNA from the *recJ*::*ble* mutant strain and the wild-type were included in the PCR reactions. The expected size for the WT *recJ* product is 2.0 kb and that for the *recJ*::*ble* product is ~1.7 kb, while a *recJ*::*km Synechocystis* mutant was expected to give a ~2.3 kb product. As shown in Figure 3.15.A, the PCR product sizes obtained confirm that the transformants are *recJ*::*km* mutants.

Having made this new recipient line, the next stage was to isolate the 'naked' *ble* cassette (*i.e.* with its own promoter that is functional in *Synechocystis* but no flanking DNA) to be used for transformation. The cassette was isolated by amplification from the pZΔES plasmid that contains the *ble* gene under the

control of the bacterial *bla* promoter, using primers listed in Appendix 3. The 0.6 kb *ble* cassette was successfully amplified as seen in Figure 3.15.B.

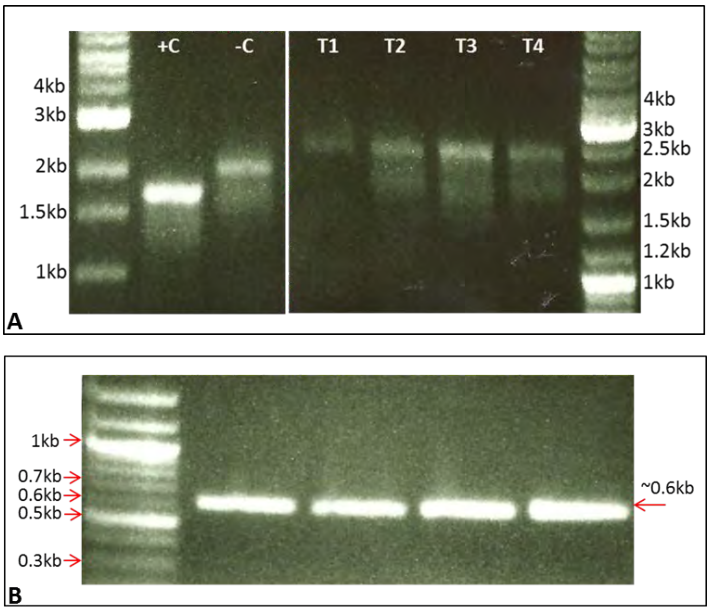


Figure 3.15: PCR confirmation for the isolation of Δ*recJ*.km strain.

(A) PCR results confirm that transformants T1, T2, T3 and T4 are Δ*recJ*.km strains (predicted band size ~2.3 kb). +C: Δ*recJ.ble* strain (~1.7 kb). –C: wild-type strain (2 kb). (B) Amplification of the *ble* cassette (~0.6 kb) from plasmid pZΔES.

3.3.5 Transformation of the Δ*rec*J.km strain with the *ble* cassette

The *ble* PCR product was purified and used for transformation into both the WT and Δ*recJ*.km strains as per methods described in section 2.4.2. Three independent transformations were performed using the Δ*recJ*.km recipient strain to ensure the isolation of independent transformants. Both positive (*i.e.* transformation with plasmid pPsb28-2::ble which contains the cassette flanked by sequence from the non-essential *Synechocystis* gene *psb28-2*) and negative (no DNA) controls were included for each of the WT and the Δ*recJ*.km strain transformations. The transformants were selected with 25 µg/ml of zeocin with putative transformants appearing after ~10 days. Whilst many colonies were obtained from the Δ*recJ*.km recipient, very few colonies were obtained from the wild type recipient strain (Figure 3.16 and Figure 3.17, respectively). The two figures also show a few colonies on the negative control plates (no DNA) from both the wild type and the Δ*recJ*.km recipient strains. This is probably due to spontaneous resistance to the antibiotic or the depletion of the antibiotics on the plates.

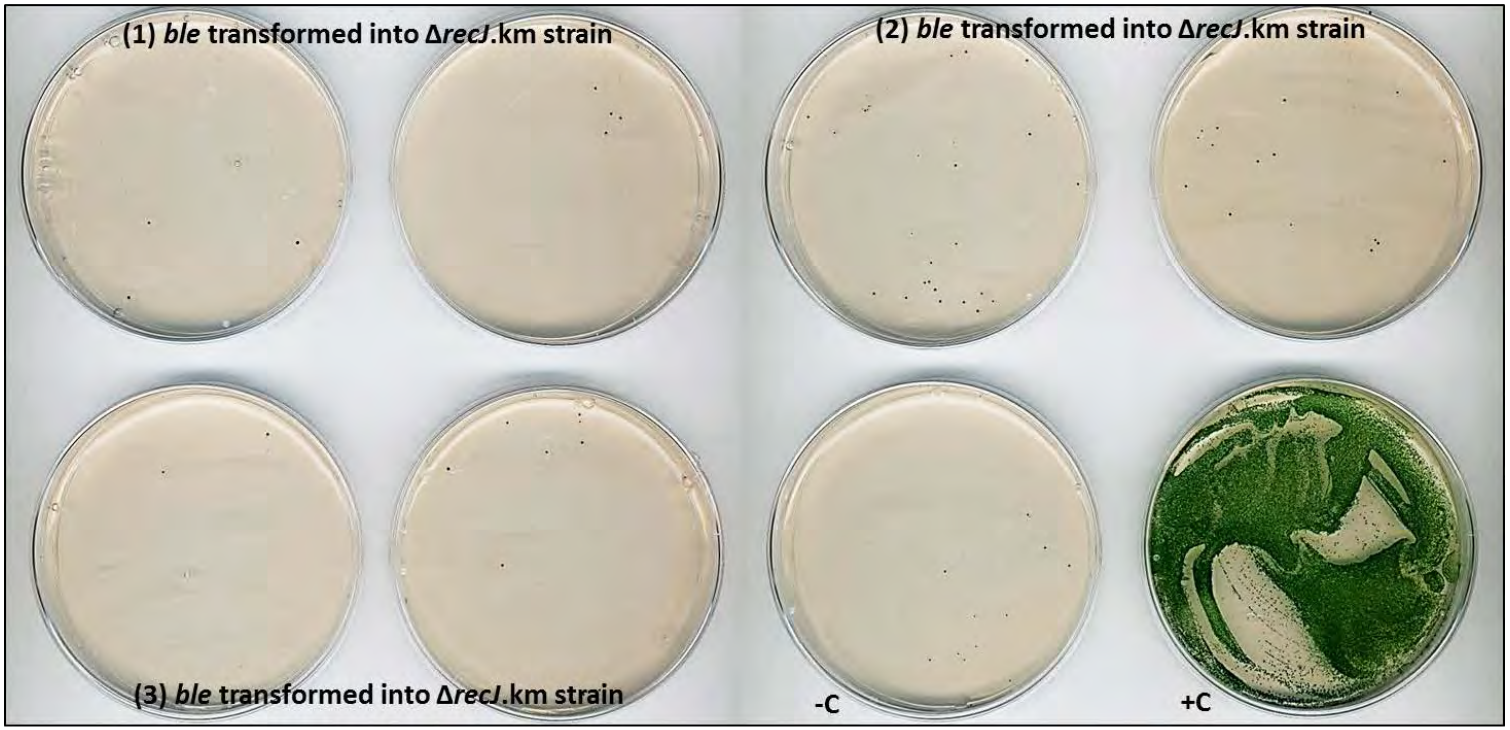


Figure 3.16: Transformants obtained from transformation of *ble* cassette in Δ*recJ*.km strain.

The figure shows transformants obtained from three independent transformations of *ble* into the Δ*recJ*.km recipient strain. Negative and positive controls are Δ*recJ*.km strain with no added DNA and Δ*recJ*.km strain transformed with a Δ*psb28*.2.ble plasmid (*psb*28.2 gene interrupted with a *ble* cassette), respectively.

Figure 3.17: Zeocin-resistant colonies obtained from transformation of the *ble* cassette in the WT 6803 strain.

Very few colonies were obtained in the WT 6803 recipient strain. Negative and positive controls are the WT strain with no added DNA and the WT strain transformed with the ΔRecJ.ble plasmid, respectively.

3.3.6 Analysis of Δ*recJ*.km *Synechocystis* transformed with the *ble* cassette.

Five transformants from each of the three independent Δ*recJ*.km strain transformations (total of 15 transformants) and five colonies from the WT transformation were re-streaked three times on selective medium to achieve homoplasmy. The colonies obtained from the transformation of the WT strain died off after the first round of re-streaking, suggesting that they were false transformants that arose due to local depletion of antibiotic on the plate. Total genomic preparations of the 14 isolated transformants from the Δ*recJ*.km strain were made according to the method described in section 2.3.3.2 and used directly for Southern blot analysis (one sample lost during preparation).

3.3.6.1 Southern blot analysis

In order to determine whether the *ble* marker gene was randomly integrated into the genome of the isolated transformants, a Southern blot was set up with all 14 lines that were derived from three independent transformation events. For this, genomic DNA prepared in section 3.3.6 was digested with *Xmn*I, an enzyme that does not cut the *ble* cassette according to methods described in section 2.3.11.2. The Southern blot was set up and a labeled probe for the *ble* DNA was prepared according to methods detailed in section 2.5.1 and 2.5.2, respectively. Figure 3.18 shows the results of the Southern blot and reveals that the *ble* cassette appears to be inserted in two distinct locations in the host genome indicated by the presence of two distinct bands in each of the transformants analysed. As with the results from the km^R cassette transformations, this result indicates that integration of foreign DNA into the *Synechocystis* genome is not random, but results in integration into specific sites even in the absence of any flanking sequences to mediate homologous recombination. The next question is whether the *ble* cassette has integrated into the 'same' site as the km^R cassette – *i.e*. the *speA*/*psbAII* locus. This is investigated in the next section.

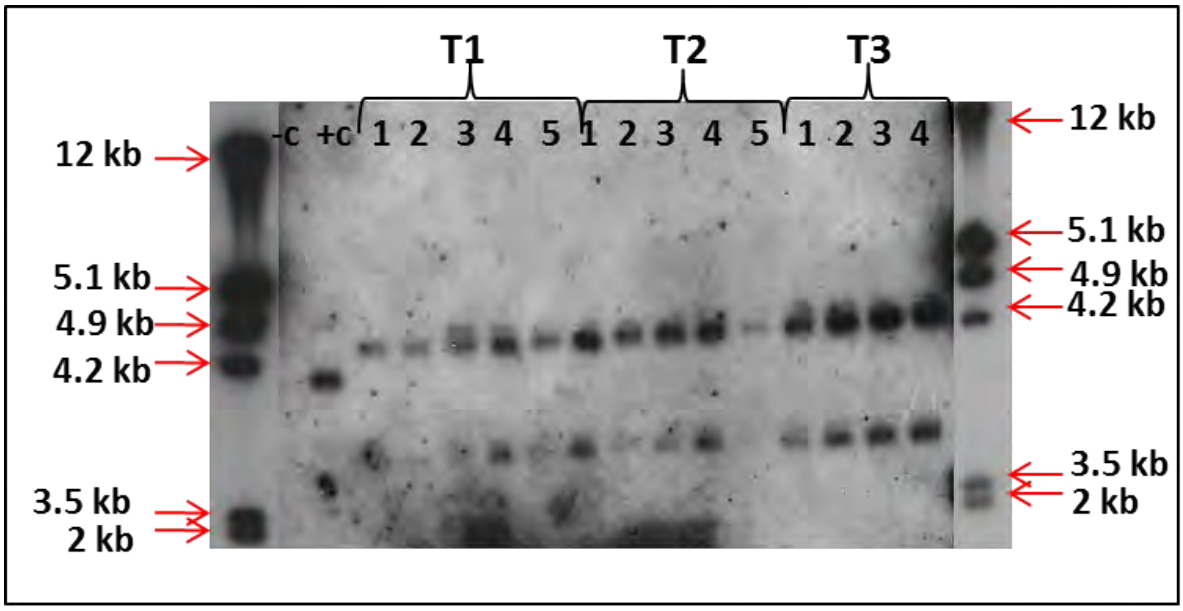


Figure 3.18: Southern blot analysis of three sets of transformants (T1-T3) from the Δ*recJ*.km strain transformed with the *ble* cassette.

All 14 transformants analysed show the same integration pattern for the *ble* selectable marker. -C: Δ*recJ*.km strain, +C; Δ*recJ*.ble strain. T1-T3 represents the three independent transformations while 1-5 represent the number of transformants from each independent transformation.

3.3.6.2 Determining whether the *ble* marker has integrated at the *speA*/*psbAII* locus.

As revealed in section 3.3.3.4, the km^R cassette appeared to be landing in the *spe*A/*psbAII* region of the genome and the integration was accompanied with a deletion of the *psbAII* locus. The same region in the new Δ*recJ*.km strain transformed with the *ble* cassette was therefore examined by PCR analysis to check for the integration of the cassette. For this PCR, three independent transformant lines were selected for analysis (since all 14 transformants showed the same integration pattern, as seen in Figure 3.18) and primers that amplify the *spe*A/*psbA*2 area of the genome (Appendix 3) were used. Controls included in the PCR reaction were WT DNA and DNA extracted from the Δ*recJ*.km recipient strain. The PCR results are presented in Figure 3.19 and show that *psbAII* region is unaffected in the transformant lines, indicating that the *ble* cassette integrates elsewhere in the genome. Further analysis needs to be done in order to identify the location at which the *ble* marker was integrated and to understand why the km^R and *ble* cassettes favoured the integration at these specific locations within the genome.

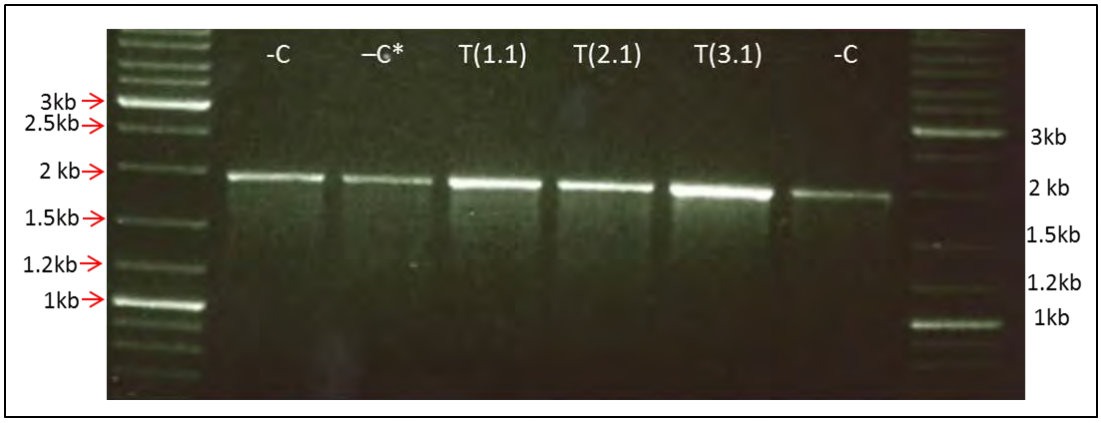


Figure 3.19: PCR analysis to investigate the itegration pattern of Δ*recJ*.km transformed with ble cassette.

Analysis of three independent transformants (T1.1, T2.1 and T3.1) show that the *ble* selectable marker had not integrated in the *spe*A.*psbA*2 region. –C: WT genomic DNA, -C*: Δ*recJ*.km recipient strain.

3.4 Conclusion and future work

The results obtained in this work provide the following picture of foreign DNA integration into the genome of *Synechocystis* 6803. When using a WT (*i.e. recJ*+) strain, successful integration requires the foreign DNA to be flanked by regions of

genomic DNA homologous to the site of integration, thereby allowing a double homologous recombination to take place. This is confirmed by the failure to obtain any true transformants for the wild type recipient strains of *Synechocystis* when either 'naked' (*i.e*. lacking any flanking sequences) km^R or *ble* selectable markers were used in the transformation.

In contrast, when the *recJ* deletion strains of *Synechocystis* are used as recipients, then transformants can be recovered (albeit at a lower efficiency) with both the naked km^R and *ble* selectable markers. We can therefore conclude that large tracts of flanking sequence [typically > 500 bp (Vermaas, 1998)] serve to protect the cassette from degradation by the RecJ exonuclease, and also provide sequences for efficient homologous recombination, but that in the absence of RecJ the cassette is not rapidly degraded, and even in the absence of extensive homology, integration of the cassette can still occur. However, both Southern blot analysis (Figures 3.6 and 3.7) and sequence analysis (Figure 3.10 and 3.13) confirmed that the integration of the km^R cassette into the genome of the Δ*recJ.*ble recipient strain was not at random locations but rather directed to a specific locus within the genome (*speA/psbAII*). On the other hand, although sequence analysis was not performed on the Δ*recJ.*km.ble transformants due to lack of time, both Southern blot analysis (Figure 3.18) along with PCR results (Figure 3.19) confirm that the *ble* marker was also not integrated randomly in the genome, but neither was it landing in the *speA/psbAII* locus. Intriguingly, the Southern blot analysis seemed to indicate integration at two specific loci. There are three possibilities as to why two bands appear on the blot; the *ble* marker has either (i) been inserted in two distinct locations "hot spots" in the genome or (ii) the DNA remains free in the cell and migrating to two locations on the gel or (iii) an experimental error was made and the *ble* marker was accidently digested with *XmaI* (that cuts in the middle of *ble*) rather than *XmnI* (which does not cut *ble*). Future sequence analysis of the transformants would reveal the answer to that question. However, whatever the case, the fact that the two selectable markers do not integrate at the same locus raises a question as to whether integration is related to the sequence of the actual selectable marker used in the transformation rather than the site of integration. For example, it has been reported that the promoter and termination sequences of kanamycin are recognised by *Synechocystis* (Vermaas, 2007). However, BLAST results of both

the km^R cassette and the *ble* selectable marker did not show any significant similarities between the sequence of the selectable markers and that of the genome of *Synechocystis*.

An interesting phenomenon reported by Ukita and Ikeda. (1996), is that the *recJ* gene product in *E. coli* is preferentially involved in the illegitimate recombination at hotspots of the genome. However, it was also reported that the illegitimate recombination at non-hotspots still took place in *recJ* mutants as direct repeat sequences were also found in the non-hotspot sites. From the previous discussion, it is also possible that the integration of both the km^R and *ble* markers is affected by the presence of repeat sequences around the site of integration. Further investigations need to be made in order to understand the integration patterns of foreign DNA into *recJ*- strains by conducting experiments aimed at testing the integration of a range of selectable markers. Nevertheless, we can conclude that using the *recJ*- strains in order to insert heterologous DNA randomly across the *Synechocystis* strain does not look feasible. Further genetic studies on illegitimate recombination will be required to clarify these points.

Chapter 4: Development of a ‘one-step’ PCR-based technique for insertion of foreign DNA into *Synechocystis* 6803

4.0 Introduction

DNA-mediated transformation in the cyanobacterium *Synechococcus* sp. strain PCC7943 signaled the beginning of genetic exchange systems in cyanobacteria (Golden et al., 1987). Today the method remains the primary means for gene transfer in unicellular cyanobacteria. There are a few species that belong to two families of unicellular cyanobacteria (*Synechococcus* spp. and *Synechocystis* spp.) that are known to be naturally transformable by exogenous DNA and therefore amenable to genetic studies (Labarre et al., 1989). Because of the ease and efficiency of the transformation procedure and the metabolic flexibility of *Synechocystis* (Kufryk et al., 2002) it has been widely used to study diverse biological processes (Koksharova, 2002). Modifications to wild type genes can be done by various means, including interruption or deletion of the whole gene, introduction of site-directed mutations, deletion of domains within a gene, and gene replacement by genes from other organisms. Such modifications of wild-type genes can be introduced effectively only in systems with homologous recombination mechanisms such as that found in *Synechocystis* (Vermaas, 1998). In all cyanobacteria for which gene transfer is feasible, the popular method of choice for targeted gene inactivation is the interruption of the cloned gene with a selectable marker (Golden et al., 1987, Vermaas, 2007).

To do this, a plasmid is constructed that can replicate in *Escherichia coli* but not in *Synechocystis* (*e.g.* pUC or pBR derivatives). This plasmid should contain *Synechocystis* sequences upstream and downstream of the gene to be deleted, with the gene itself being replaced by a selectable marker of choice. Two crossover events take place, followed by a selective pressure that selects for the acquisition of the foreign DNA (*e.g.* selection for antibiotic resistance) resulting in the recovery of transformants carrying the modified genotype (Vermaas, 1998). An illustration of this method is found in Figure 4.1.

Figure removed for copyright reasons

Figure 4.1: A schematic representation of double homologous recombination in *Synechocystis.*

Double homologous recombination between the cyanobacterial genome and an introduced plasmid carrying *Synechocystis* sequences identical to those upstream and downstream of a cyanobacterial target gene (upstream and downstream elements are indicated by open and closed ovals). The homologous recombination results in the replacement of the chromosomal segment with a selectable marker.

4.1 Literature review

4.1.1 Current methods for gene replacement in *Synechocystis*

In *Synechocystis*, transformation by exogenous DNA is done using either shuttle plasmids that can replicate independently of the cyanobacterial genome or by integration of the transforming DNA into specific loci in the genome via homologous recombination (Chauvat et al., 1986, Grigorieva and Shestakov, 1982). The latter method is particularly valuable as a reverse-genetics tool for modifying endogenous genes – either by creating null (or 'knockout') mutants by gene disruption or by introducing site-directed changes into the gene (Koksharova, 2002). Whilst site-directed variants can also be introduced in shuttle vectors, this approach suffers from the problem of gene dosage effects if the plasmid is present in multiple copies, and the necessity of analysing the site-directed change against a background of endogenous gene activity from the wild-type gene present in the genome (Zorin et al., 2005). The traditional construction of knockout mutants in *Synechocystis* is a time consuming process (Lesic and Rahme, 2008) as it required several cloning steps. As shown in Figure 4.2, these steps are: i) isolation of genomic DNA and amplification of the target gene by PCR amplification; ii) cloning the PCR product into a restriction enzyme (RE)-cut plasmid vector; iii) isolation of a DNA fragment carrying an antibiotic-resistance marker; iv) cloning of the marker into the plasmid at a restriction site such that it disrupts the target gene; v) preparation of the final plasmid and checking by restriction digest for the correct insertion and orientation of the marker. Faster methods of creating plasmid vectors that avoid RE-based cloning have been developed recently. These methods include sequence- and ligation-independent cloning (SLIC) (Li and Elledge, 2007), Gibson assembly (Gibson et al., 2009), circular polymerase extension cloning (CPEC) (Quan and Tian, 2011), fusion PCR (Schwarz et al., 2013), Uracil-Specific Excision Reagent (USER) (Geu-Flores et al., 2007) and the Golden Gate method (Engler and Marillonnet, 2011).

The common feature between these methods is the use of multiple PCR products with overhangs that undergo special treatment to fuse the fragments to form the plasmid used for transformation. Although these methods are considered faster than traditional cloning methods, they still suffer some limitations as they require the amplification and assembly of multiple PCR products to create the gene knockout plasmid. Scientists were therefore encouraged to develop a more time-efficient mutagenesis method that does not require cloning or the need of multiple PCR products. As shown in Figure 4.2, a simple strategy is to generate a PCR product directly from the selectable marker using 5'-extended primers that introduce short flanking regions with homology to the gene to be deleted. This "one-step PCR-based method" for gene deletion thus requires only a single PCR reaction, with no genomic isolation or plasmid cloning steps, and would therefore be considerably faster than the traditional method. Furthermore, the site of insertion and orientation of the marker is not constrained by the availability of suitable RE sites, and therefore the marker can be targeted to any location in the gene. The one-step PCR-based method described above has already been described in *E. coli* (Datsenko and Wanner, 2000), *Aspergillus nidulans* (Chaveroche et al., 2000), *Yersinia* (Derbise et al., 2003, Lesic et al., 2004), *Salmonella* (Husseiny and Hensel, 2005, Beloin et al., 2003), *Shigella* (Beloin et al., 2003), *P. aeruginosa* (Lesic and Rahme, 2008) and *Serratia* (Rossi et al., 2003). However, to the best of my knowledge, no such reports exist for *Synechocystis*, or other cyanobacteria. The closest work developed for quick gene knockouts in *Synechocystis* that is independent of cloning, is the technique described by Taroncher-Oldenburg and Stephanopoulos. (2000), who used a PCR-based method for the construction of targeted gene disruptions and gene fusions in *Synechocystis* using regions of homology with the target gene. The proposed method, called long flanking homology polymerase chain reaction (LFH-PCR), involves creating a conversion cassette of the deleted gene using two separate PCR reactions to generate products with overlapping sequences that can be used in another PCR to generate a product containing the gene interrupted with the selectable marker. Although this method is faster than creating plasmid constructs, it still involves multiple PCR steps.

This chapter describes the development of the one-step PCR method for generating gene knockouts in *Synechocystis* that should greatly accelerate the

reverse-genetic analysis of any gene in this model cyanobacterium. However, a key aspect of developing such a method is identifying the minimum length of the flanking sequences required for homologous recombination.

Figure removed for copyright reasons

Figure 4.2: Comparison between the traditional and one-step PCR methods in creating gene knockouts in *Synechocystis*.

A comparison between the "traditional method" (left) and the proposed quicker "one-step PCR" (right) for creating gene knockouts in *Synechocystis.* Modified from (Vermaas, 2007).

4.1.2 The minimum length for homologous recombination

Transformation of *Synechocystis* with either circular or linear DNA (*e.g.* PCR products) requires large tracts of DNA sequence flanking the transgenic DNA to be introduced, such that the two homologous recombination events can occur efficiently. Typically, the advice is to ensure that each flanking DNA region is >500 bp. This is considered particularly important when using a linear construct, possibly as a consequence of exonuclease activity inside or outside the cell degrading the ends of the DNA (Vermaas, 1998).

The minimum length of flanking sequences required for efficient insertion of the DNA in *Synechocystis* has not been systematically studied, with most researchers simply ensuring that each of the flanking sequences are typically >500 bp in length. However, Zang et al. (2007), reported that transformation was possible with only 200 bp of flanking sequence on one side (1300 bp on the other – a "1300/200" arrangement). Furthermore, the same group showed that a "400/400" arrangement also yielded transformants. Similarly, Ikeuchi and Tabata. (2001), showed that 400 bp, but not less, was enough for a double homologous recombination in *Synechocystis*. Labarre and collegues. (1989), earlier reported success with "900/800" but very few transformants with "900/300" and no transformants with "900/0" while Vermaas. (1998), concluded that efficient double homologous recombination between the genomic DNA of the organism and

introduced DNA requires at least 100 to 400 bp sequence identity in both recombination domains. Initial studies in Saul Purton's lab by Edward Burton (an undergraduate project student) showed that a "186/171" was sufficient for insertion of a kanamycin resistance marker at the *psbN* locus. It is reported that the *psbN* gene that encodes a photosystem II associated protein PSII-N (Kaneko et al., 1996) is not essential for photosynthetic growth and is fully dispensable (Vermaas, 1998). Consequently, in the work described in this chapter, attempts have been made to use a series of different PCR products derived from the same *psbN*::km^R knockout construct to determine the minimum size of homologous flanking regions of linear dsDNA that is required for successful transformation in *Synechocystis* via double homologous recombination (Figure 4.3). This information is used to determine if we can ultimately target foreign DNA into a specific locus of the *Synechocystis* genome by simply transforming with a PCR product in which only the necessary flanking sequences are added to the foreign DNA using "long tail" primers that are homologous to the kanamycin resistance cassette at their 3' ends and homologous to the target gene at their 5' ends as illustrated in Figure 4.2.

Furthermore, since the transformation efficiency drops markedly for DNA with shorter homologous segments (Ikeuchi and Tabata, 2001), it was therefore essential to identify the shortest homologous recombination fragment necessary for efficient transformation to occur in *Synechocystis.*

4.2 Chapter aims

The aim of the work undertaken in this chapter was to develop a rapid method for creating gene knockouts in *Synechocystis* by:

- Identification of the shortest size of flanking regions necessary for efficient double homologous recombination transformation to occur in *Synechocystis.*
- To use this information to generate gene knockouts in *Synechocystis* using a quick one-step PCR-based method that is independent of a transforming plasmid or the fusion of multiple PCR products.

4.3 Results and Discussion

4.3.1 Identification of the minimum flanking length necessary for transformation

4.3.1.1 Generation of test PCR products with different flanking lengths

For this work, plasmid pUC.psbN::km^R (Appendix 1) was used as a template for the generation of a series of PCR products containing a *psbN*::km^R knockout with flanking regions ranging from 1000 bp down to 30 bp as illustrated in Figure 4.3. The primers used in this work are found in (Appendix 3).

The PCR was carried out as described in section (2.3.5) and an aliquot of the products was analysed by agarose gel electrophoresis. The expected sizes of the PCR products (~3.1 kb - 1.3 kb) were confirmed by comparison to a 1 kb ladder (Figure 4.4). PCR products were purified using the QIAquick PCR purification Kit (QIAGEN), according to the manufacturer's instructions and treated with the restriction enzyme *Dpn*I, which cuts only methylated DNA (at the sequence: G^{me}ATC). Since the PCR products are unmethylated, but the pUC.psbN::km^R plasmid used as the PCR template is methylated since it was produced in *E. coli* (which contains the *dam* enzyme that methylates GATC), then this treatment avoids the possibility of trace amounts of the plasmid contaminating the PCR product and giving rise to transformant colonies.

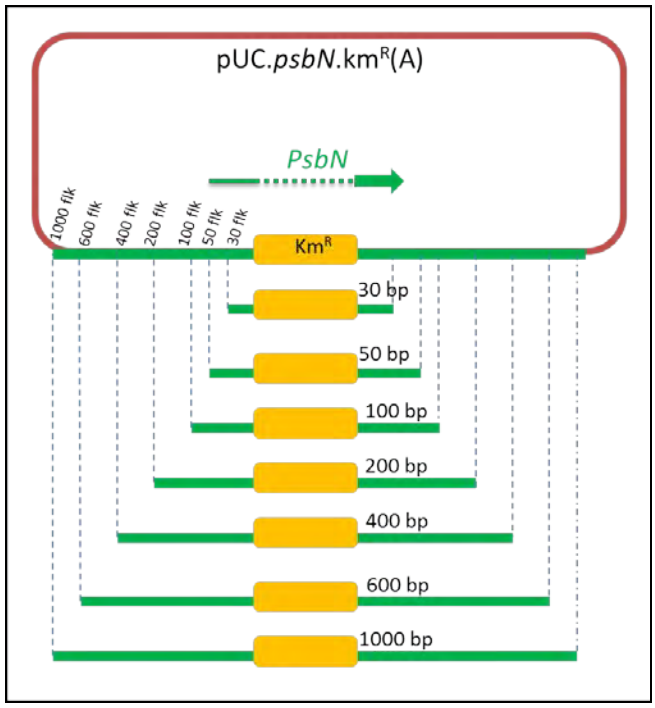


Figure 4.3: A schematic representation of the PCR products with different flank sizes.

Utilizing pUC.psbN::km^R for the generation of *Synechocystis psbN* knockouts with different flanking sizes. The *psbN* gene is interrupted by the insertion of a kanamycin resistance cassette. (1000flk -30flk) represents primers used to generate PCR producted with variouse flank sizes.

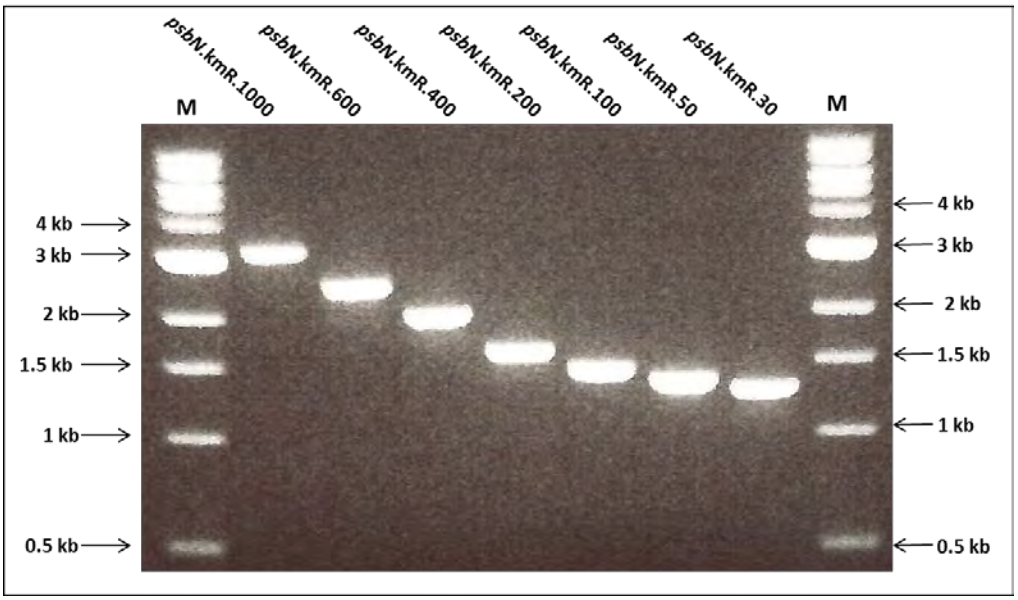


Figure 4.4: PCR amplification of *psbN*::km^R knockouts with different flanking sizes.

The amplification of PCR products derived from the same *psbN*::Km^R knockout construct with different flanking regions ranging from 30 bp to 1000 bp. M represents the 1 kb marker.

4.3.1.2 Obtaining *Synechocystis* transformants using the PCR products

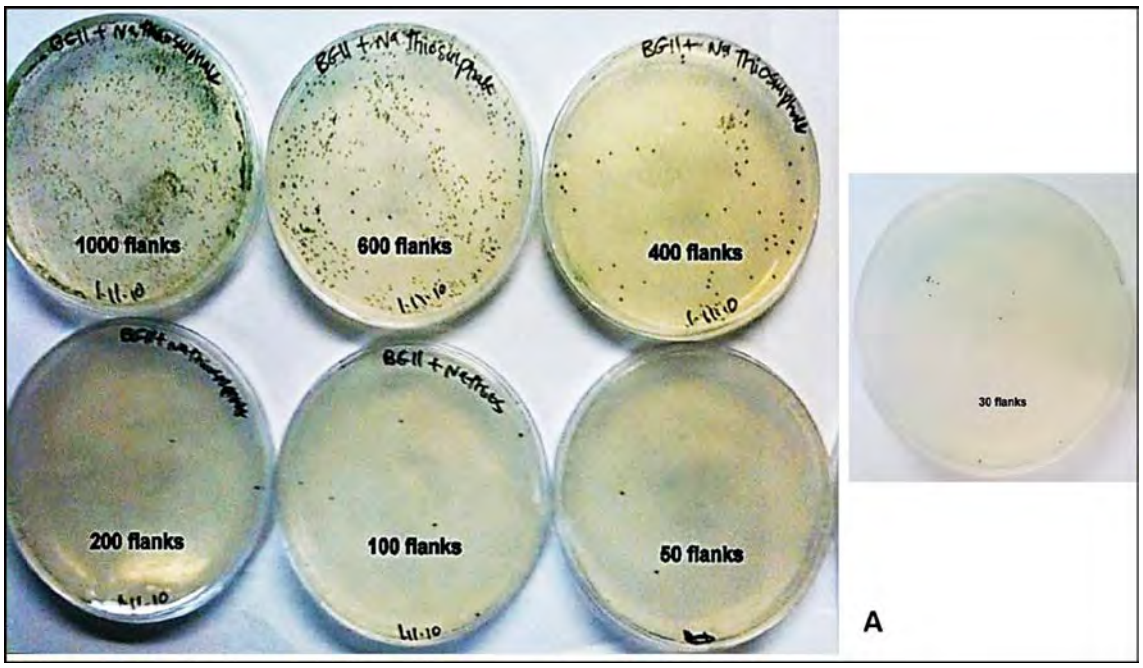
The PCR products (1–5 µg) obtained in the previous section were used for transformation into *Synechocystis* as described in section 2.4.2. According to (Zang et al., 2007), seven times more transformants survive when transformed cells are grown in medium containing no antibiotics for at least one day prior to being transferred to a medium containing the selective antibiotic. Therefore, the transformed cells were allowed to recover for ~48 hours before being overlaid with soft agar (0.6 %) supplemented with 50 µg/ml kanamycin.

The transformation was successful and putative colonies appeared on the plates within seven to ten days post antibiotic selection. The efficiency of the transformation was found to be directly proportional to the length of the flanking regions in the PCR products, with higher transformation rates obtained with PCR products with longer flanking elements (Figure 4.5.A). Figure 4.5.B shows the

average colony counts of *Synechocystis* transformed with the different PCR products. The average represents counts from three independent transformations, although error bars are not included in the plot as the three transformations were not carried out under identical conditions (variable DNA concentrations and growth stage at time of transformation). However, the same trend of a decrease in colony counts with decreased flank size down to a minimum of 200 bp was observed in all three transformations. Conversely, the colony counts for products with <200 bp varied and did not follow any trend, such that in some cases, slightly more transformants were obtained with PCR products with shorter flanks (Figure 4.5). It can be concluded that although it is possible to obtain transformants using PCR products with short flanks (<200 bp), the efficiency of transformation is very low.

4.3.1.3 Analysis of the transformants

Individual colonies were picked and re-streaked onto fresh medium supplemented with kanamycin. This was repeated three times to ensure that the cells achieve homoplasmy before isolating genomic DNA as described in section (2.3.3.1). The extracted DNA was used as a template in a PCR reaction using the primers named "1000 flanks" (Appendix 3) – *i.e.* outside of the region of the transforming DNAs to check for the correct integration into the *psbN* locus. PCR conditions were carried out as described in section (4.3.1.1) and results presented in Figure 4.6.A show the successful isolation of transfromants carrying a *psbN*::km^R knockout. Controls were included for band size comparisons. Both a plasmid carrying a WT *psbN* gene and genomic WT DNA were used as negative controls while pUC.psbN::km^R was used as a positive control. A band shift from 1.9 kb (representing the WT *psbN* gene) to a band size of 3.1 kb (*psbN* gene including the km^R cassette) confirms the isolation of *psbN*::km^R knockouts (Figure 4.6.B).



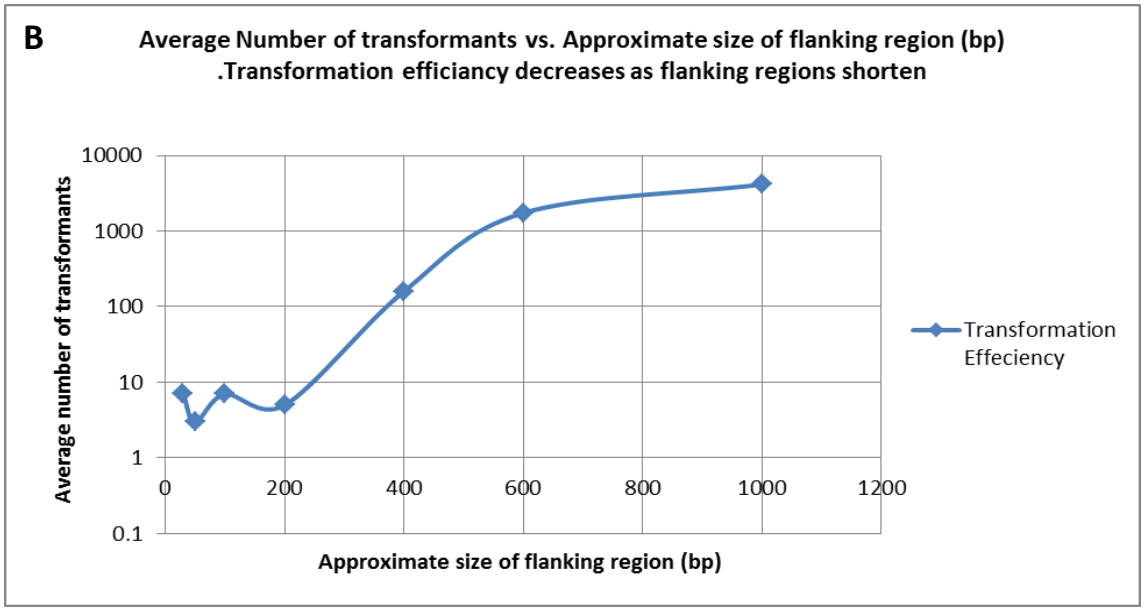


Figure 4.5: Transformants generated using the traditional method.

(A) *Synechocystis* transformants obtained upon transformation with *psbN*::km^R PCR products with 1000 bp to 30 bp flanks on either side. (B) Transformation efficiency represented by the number of transformants decreases with reduced flank size.

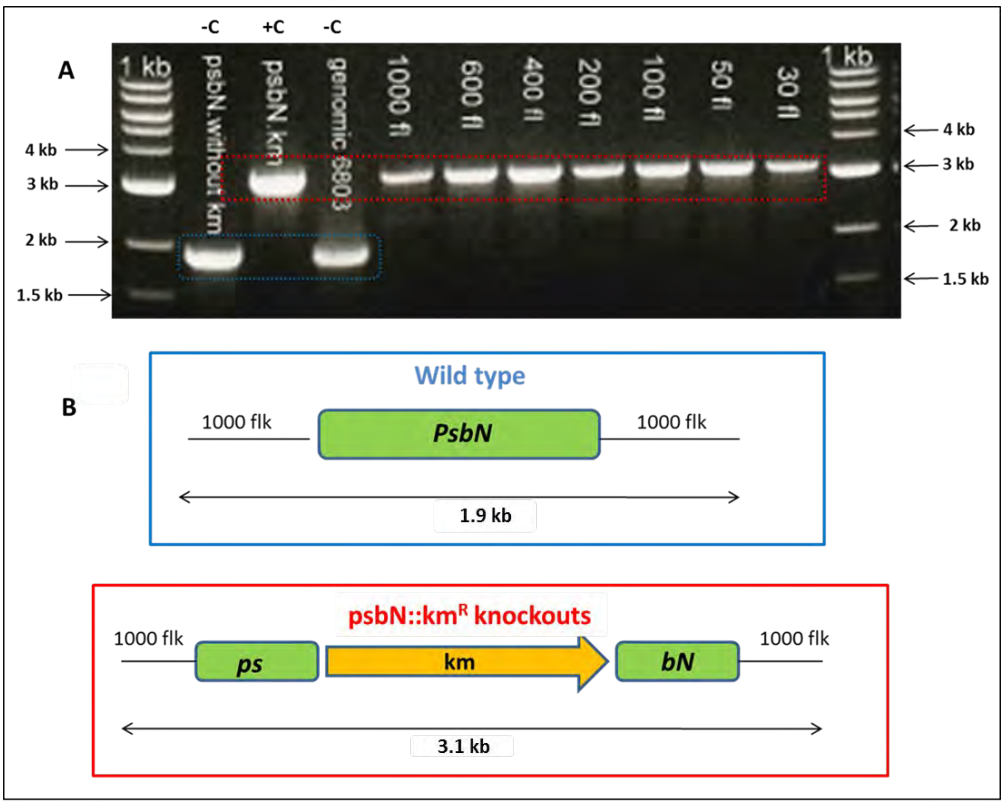


Figure 4.6: PCR results confirming isolation of *psbN*::km^R knockouts using the different PCR products.

(A) PCR results confirming the isolation of *psbN*::km^R knockouts in *Synechocystis* even with the shortest flank (30 bp). (B) WT size = 1.9 kb (represented within the blue rectangle), transformants size = 3.1 kb (represented within the red rectangle). Controls in order: (-C): plasmid carrying WT *psbN* gene (plasmid A without km^R), (+C): pUC.psbN::km^R, (-C): WT genomic DNA.

Despite several published reports on the failure to produce successful gene knockouts in *Synechocystis* by means of double homologous recombination using flanking sequence less than 200 bp (section 4.1.2), this current study establishes the possibility of creating targeted gene knockouts using PCR products containing as little as 30 bp flanking regions. Nevertheless, the efficiency of the transformation did drop markedly with decreased flanking size. However, findings presented in this section opened the door to investigate the possibility of creating gene knockouts using a quick one-step PCR-based method as described in the following section.

4.3.2 Testing the one-step PCR-based method for generating gene knockouts in *Synechocystis*

Work done in this section was aimed at investigating the feasibility of targeting foreign DNA into a specific locus of the *Synechocystis* genome by simply

transforming *Synechocystis* with a PCR product in which the necessary flanking sequences are added to the foreign DNA using “long tail” primers that are homologous to a resistance cassette at their 3' ends and homologous to the target gene at their 5' ends. However, since the maximum length of primer that can be realistically synthesised with reasonable cost is ~100 bases in length, and if ~20 bases is typically needed for priming, then this means that the maximum flanking sequences that can be created using this one-step PCR-based approach is “80/80”. According to the results from section 4.3.1 this should be feasible.

4.3.2.1 Generation of PCR products using the one-step PCR method

PCR was employed to amplify the kanamycin resistance cassette directly from the plasmid pUC4K (Appendix 1) and recreate a *psbN*::km^R product using long-tail primers. The primers were designed such that 22 bp were homologous to the resistance cassette at their 3' end and short flanks (30 bp, 50 bp and 80 bp, respectively) were homologous to the target gene (*psbN*) at their 5' ends as illustrated in Figure 4.7.A.

PCR results presented in Figure 4.7.B show the successful amplification of *psbN*::km^R products in a single PCR step using the long-tail primers. Once the correct PCR product size (~1.3 kb) was confirmed, the PCR product was purified and used in transformation of WT *Synechocystis* recipient cells. *Dpn*I treatment was not necessary in this case; since the pUC4K used as a template to generate *psbN*::km^R PCR products does not contain genomic *Synechocystis* DNA and therefore there is no risk of generating false positive transformants. Sequences of the long-tail primers used are listed in Appendix 3.

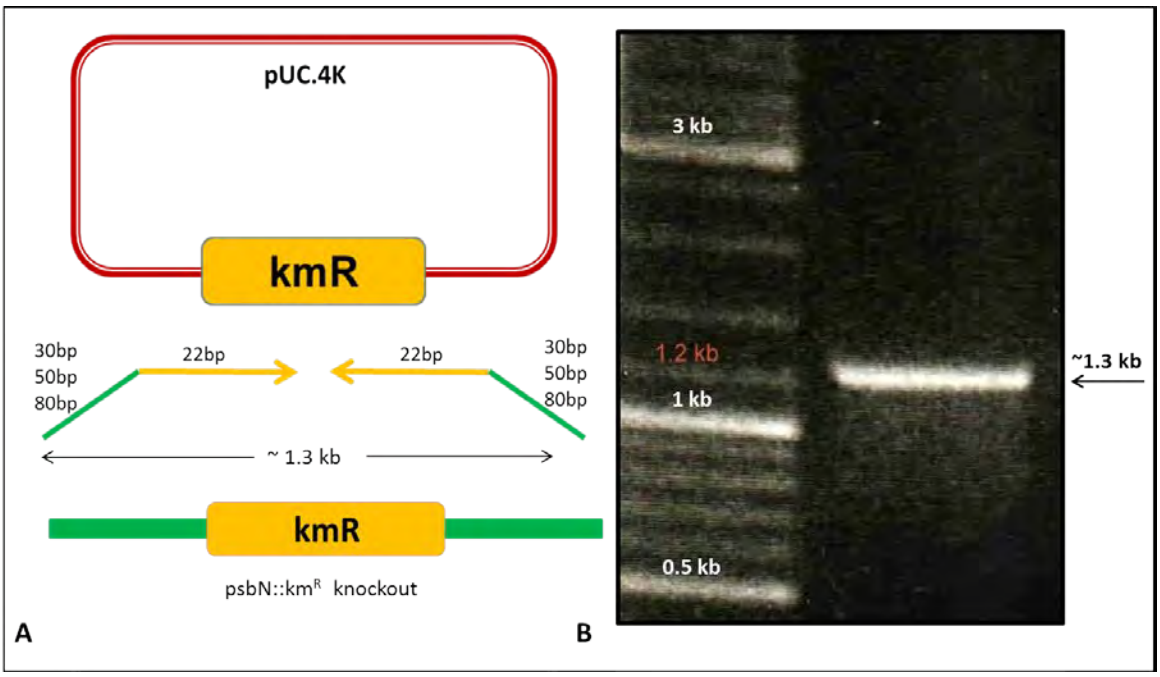


Figure 4.7: PCR amplification of *psbN*::km^R knockouts using one-step PCR method.

(A) Schematic representation illustrating the amplification of the *psbN*::km^R knockout with short flanks (less than 100 bp) using one-step PCR. (B) The successful amplification of *psbN*::km^R knockouts in a single PCR step. PCR product shown in the figure is with 30 bp flanks homology.

4.3.2.2 Transformation of one-step PCR products in *Synechocystis*

Approximately 1-5 µg of the PCR products obtained in section 4.3.2.1 were used for transformation into wild type *Synechocystis* as per methods described in section 2.4.2. Many colonies were recovered upon transformation of the PCR products within seven to ten days post antibiotic selection using 50 µg/ml of km. Surprisingly, thousands of colonies were obtained using this method for PCR products of *psbN*::km^R knockouts with 30 bp homology to the target gene (hereafter LT30) while those obtained from both *psbN*::km^R PCR products with 50 bp and 80 bp homology to the target gene (hereafter LT50 and LT80, respectively) showed a few hundred colonies, and four colonies, respectively as seen in Figure 4.8. This is contrary to what was expected and observed earlier where fewer transformants were obtained when shorter regions of homology to the target gene were utilised in the transformation. Furthermore, the number of transformants obtained with PCR products with 30 bp and 50 bp homology obtained from the one-step PCR method presented here was much higher compared to the number of colonies obtained with the same size flanks using derived by PCR from the original pUC.psbN::Km^R plasmid (Figure 4.5.A). This

observation raised doubts as to whether the transformants were actually obtained as a result of double homologous recombination or via other means. The transformants were probably not due to spontaneous resistance to kanamycin since no colonies were observed on the negative control ('no DNA') plate.

4.3.2.3 Analysis of *psbN*::km^R knockouts generated using the one-step PCR method

Thirty independent transformants obtained using either the LT30 and LT50 PCR products, and the four transformants obtained using the LT80 PCR product were re-streaked on BG11 plates supplemented with kanamycin added to a final concentration of 50 µg/ml. From these plates, colonies were further restreaked twice, to ensure that all copies of the polyploid chromosome contained the introduced DNA. Genomic DNA of the transformants was prepared as described in section 2.3.3.1. The extracted DNA was used as a template in a PCR reaction as before (section 4.3.1.1) using primers of 1000 flanks to check for the correct integration into the *psbN* locus. PCR results presented in Figure 4.9 show that none of the tested transformants gave a positive result for a *psbN*::km^R knockout. The same positive and negative controls were included in the PCR reactions as those used in section 4.3.1.3 including an extra positive control (a positive *psbN*::km^R knockout obtained in the previous section). To confirm this result, the transformation was repeated using the same methods described above for the LT30, LT50 and LT80 transformants and out of hundreds of colonies obtained from the transformations using LT30 or LT50 PCR products (no colonies appeared using the LT80 PCR product) five colonies of each transformation were screened for targeted gene replacement but again, all ten transformants screened showed a wild type band size of 1.9 kb. Figure 4.9 shows PCR results of three representatives of each transformation type.

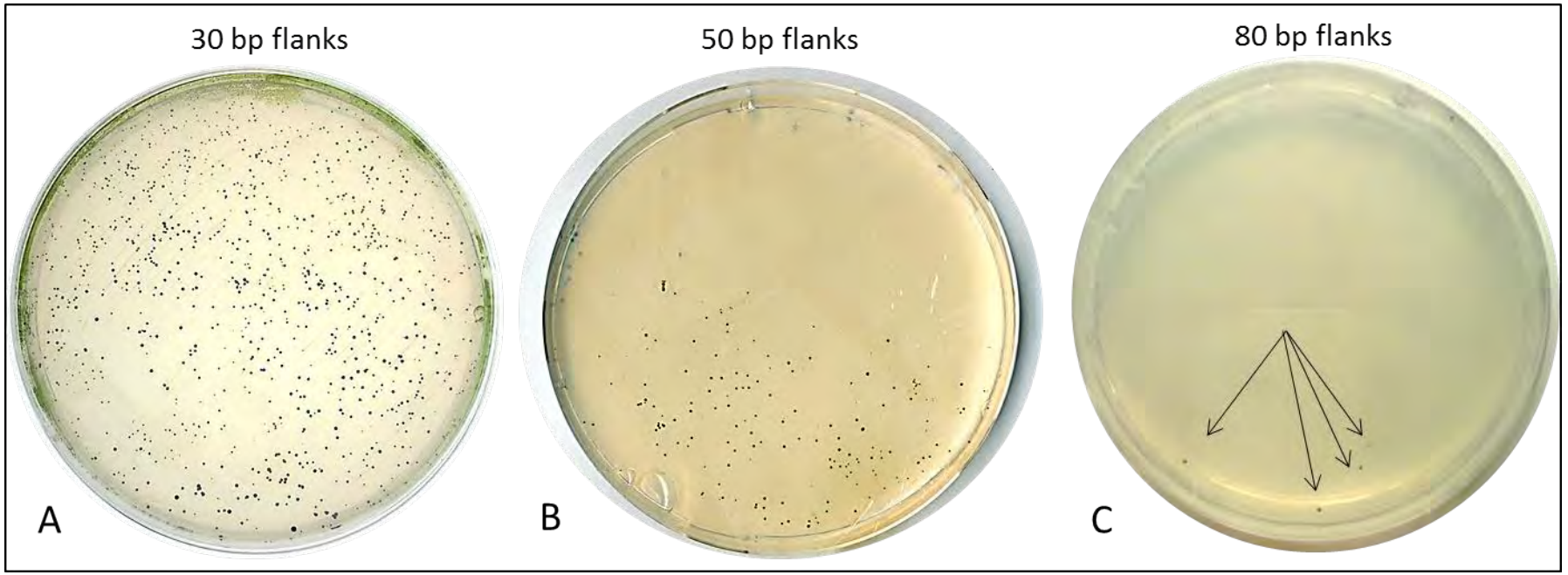


Figure 4.8: Transformants obtained by one-step PCR using short flank PCR products.

The number of transformants obtained decreased with increased flank size contrary to what was expected. Selection was on BG11 plates supplemented with 50 µg/ml of km. The 30-80 bp flanks represents transformants obtained with PCR products with 30 bp, 50 bp and 80 bp flanks on either end of the km^R cassette.

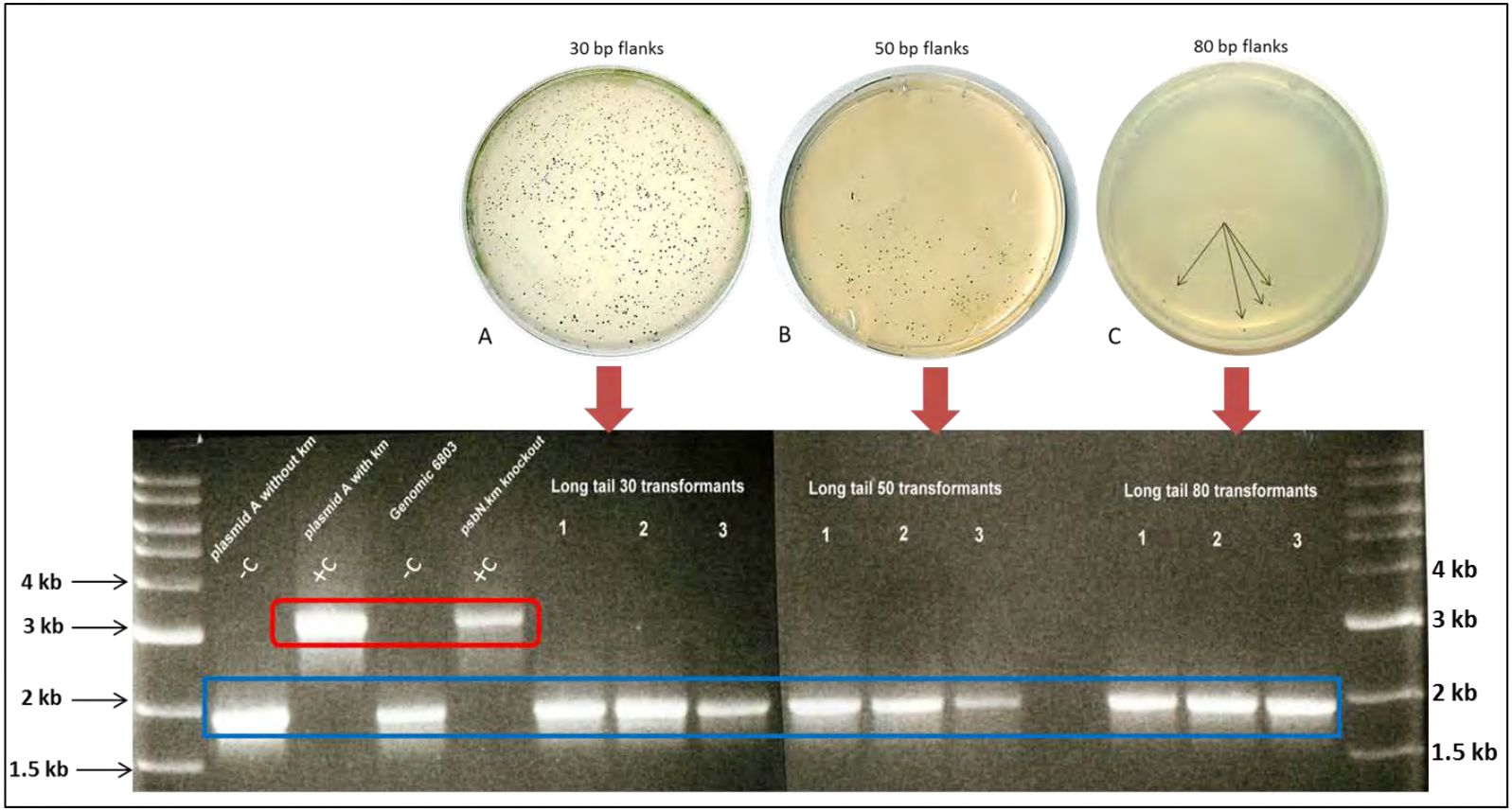


Figure 4.9: PCR results of transformants obtained using products from a one-step PCR.

Top figure shows *Synechocystis* transformants generated from PCR products with 30 bp, 50 bp and 80 bp flanks using the one-step PCR method (A, B and C respectively). Bottom figure shows PCR results confirming the absence of *psbN*::km^R knockouts in any of the transformants. Wild type bands are enclosed in a blue rectangle (1.9 kb) whereas positive controls are enclosed in a red rectangle (3.1 kb). Controls are indicated within the figure.

In an attempt to understand the reason behind the high numbers of antibiotic-resistant transformants that are not *psbN* gene disruptions obtained from the one-step PCR-based method, predictions were made on the possible events that might have taken place during transformation. There are three possible ways in which km resistant transformants can be obtained: firstly, the one-step PCR product integrated via double homologous recombination (HR) resulting in *psbN* knockout. Secondly, the PCR product inserted into the genome by means of non-homologous recombination (NHR) and finally only the km^R cassette was integrated into the genome without the short flanks possibly through pUC4K contamination.

In order to identify which of the three events took place; a PCR was carried out on the genomic extract of a representative LT30 transformant using *psbN* locus-specific primers (1000 flank primers), PCR product-specific primers (30 flank primers) and kanamycin-specific primers (km primers). The primer positions and predicted PCR outcomes are illustrated in Figure 4.10.A. The PCR was carried out under the same conditions described in section 4.3.1.1. The results of the PCR represented in figure 4.10.B show a positive result for both the PCR product-specific primers and the km^R primers (same bands as the positive control). This result confirms the integration of the entire PCR product (*i.e*. the km^R cassette along with the 30 bp *psbN* flanks on either side of the cassette) and rules out the possibility of contamination with traces of the pUC4K plasmid. However, as seen earlier, the results show a negative result for the *psbN* locus-specific primers (1000 flank primers) suggesting that the PCR products were integrated elsewhere into the *Synechocystis* genome via non-homologous recombination.

In order to understand the reason behind the different integration patterns (HR *versus* NHR) between the PCR products of *psbN* gene knockouts obtained from the PCR products generated from the pUC.psbN.Km^R plasmid and the one-step PCR products (respectively), attempts were made, to identify the factors that may have led to the different integration patterns.

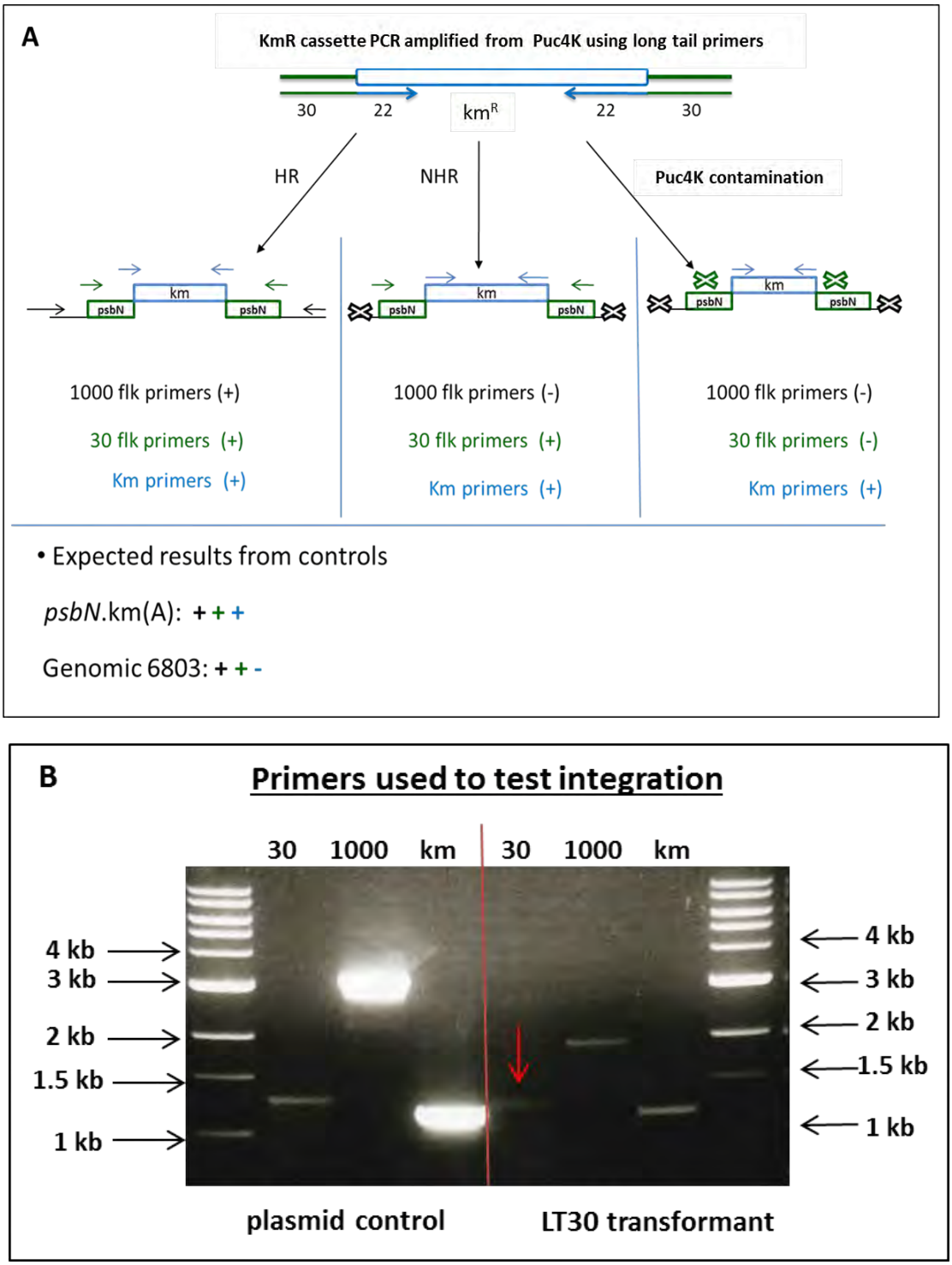


Figure 4.10: Analysis of transformants obtained by one-step PCR with short flanks.

(A) Illustration showing PCR set up to investigate scenarios of possible integration of the PCR product with 30 bp flanks obtained from one-step PCR method (B) PCR results confirm the non-homologous integration. Primers 30, 1000 and km were used to test integration of the entire PCR product (LT30), integration of PCR product in *psbN* locus and insertion of the km^R cassette respectively. The plasmid control used was pUC.psbN::km^R.

4.3.2.4 Variability in growth and transformation conditions

It may be possible that the integration patterns of the PCR products of *psbN*::km^R knockout obtained from the two methods showed variability due to factors related to the conditions in which the transformation was carried out. Therefore, PCR products with flanking regions ranging from 1000 bp to 30 bp (made by the traditional method – Figure 4.4) and an LT30 PCR product of the *psbN*::km^R knockout (from the one-step PCR method) were transformed into *Synechocystis* under the same conditions. Transformation was carried out as described in section 2.4.2 and using 2 µg of DNA and a recipient cell concentration of 4x10^8 cells/ml. The transformants were allowed to recover for 48 hours before overlaying them with kanamycin added to a final concentration of 50 µg/ml on the plates.

Putative colonies appeared on the plates from all transformations and showed the expected trend where fewer colonies were obtained with PCR products with shorter flanks. Interestingly, even the transformation with the PCR product produced via the one-step PCR method gave only a single colony, which is contrary to previous results where hundreds of colonies appeared from the PCR product with 30 bp of homology (Figure 4.9). This result in itself was a good indication that the one-step PCR product had in fact been integrated via double homologous recombination. To confirm this, PCR analysis was carried out including the same positive and negative controls used in section 4.3.2.3. Results presented in Figure 4.11 confirm the isolation of *psbN*::km^R knockouts with PCR products obtained from both methods.

To compare the kanamycin resistance of the transformant obtained using the one-step PCR method (LT30) obtained in this experiment (*psbN* knockout) to those (non-*psbN* knockouts) isolated using the same method obtained in section 4.3.2.3, 'spot tests' were carried out as described in section 2.2.6. Despite the fact that the LT30, LT50 and LT80 (NHR) transformants had previously survived three rounds of re-streaking on BG11 plates supplemented with 50 µg/ml of kanamycin, spot tests of the same isolated transformants (two months later) failed to grow under the same concentration of kanamycin, indicating loss of resistance, probably due to loss of the inserted km^R cassette (Figure 4.12.A). This also suggests the possibility of the one-step PCR products being linked to each other in circular structures (concatamers), independent of the host genome, that

are ultimately lost upon successive replication. On the other hand, the true *psbN*::km^R knockout obtained in Figure 4.11 (LT30) showed resistance to even higher concentrations of km (200 µg/ml) for several generations (Figure 4.12.B), further confirming proper integration of the km^R cassette.

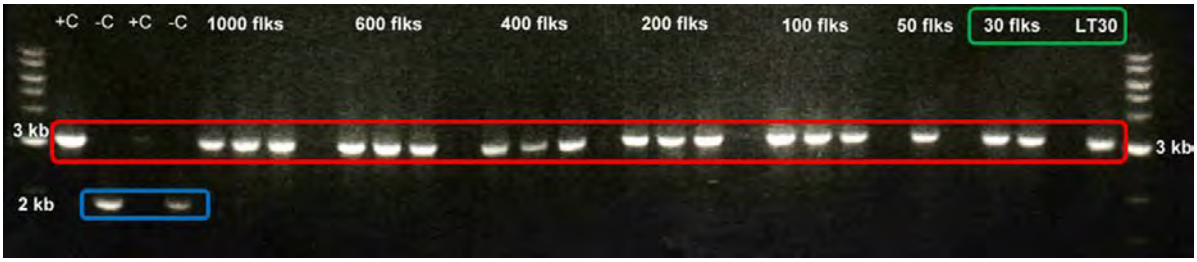


Figure 4.11: Transformants obtained with the traditional and one-step PCR methods.

PCR results confirm the isolation of *psbN*::km^R knockouts (~3 kb) even with PCR products with the shortest flanks (enclosed in green rectangle) using both the traditional and one-step PCR methods. +C: plasmid A and a *psbN*::km^R knockout, respectively. – C's: Plasmid A lacking the km^R cassette and genomic DNA respectively.

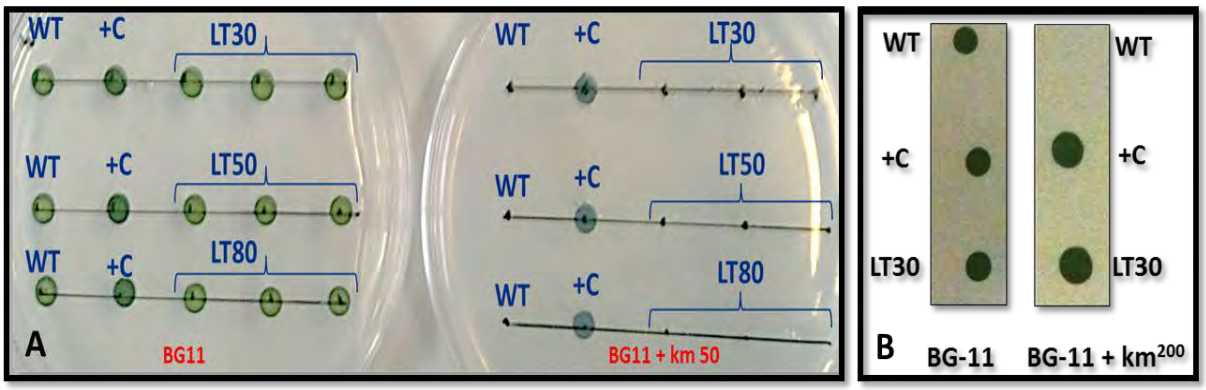


Figure 4.12: Spot test showing tolerance of *Synechocystis psbN* knockouts to kanamycin.

(A) LT30, LT50 and LT80 transformants lose resistance to kanamycin after many generations. (B) *Synechocystis psbN* knockouts obtained from one-step PCR showing resistance to high concentrations of kanamycin (200 µg/ml). -C: WT cells, +C: Δ*recJ*::km^R *Synechocystis* strains.

To summarise, the transformants of PCR products with short flanks obtained from the traditional method normally resulted in the isolation of targeted gene knockouts (with very few exceptions) while on the other hand, transformants of PCR products with short flanks obtained from the one-step PCR method rarely resulted in the isolation of targeted gene knockouts (an exception was the single transformant shown in Figure 4.11).

From these findings, it was hypothesised that non-homologous recombination was significantly more likely than homologous recombination with PCR products obtained from the one-step PCR method. It is also possible that *Synechocystis* likely has a recombination system that could insert such a cassette in "hot spots" in the genome (as seen in chapter 3), probably when the homologous sequences flanking the sides of the cassette were not identified. This hypothesis might explain the trend obtained in Figure 4.8 where PCR products with shorter flanks resulted in higher number of transformants possibly as a consequence of the PCR product with very short flanks (30 bp) being identified as a "naked" kanamycin cassette thus leading to its integration at these "hot spots". Homologous recombination, on the other hand, was significantly more likely than non-homologous recombination with PCR products from the traditional method for reasons that are identified later in this chapter. This hypothesis is depicted in Figure 4.13.

The following sections describe attempts to improve the ratio of homologous to non-homologous recombination events taking place in *Synechocystis* in order to create a more efficient and reproducible one-step PCR method.

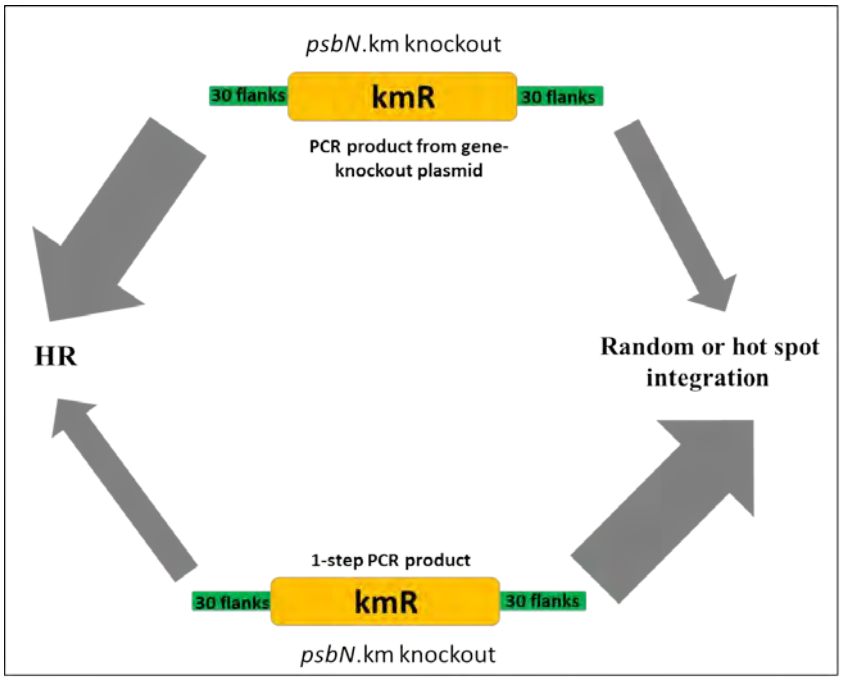


Figure 4.13: Possible integration patterns of PCR products with short flanks obtained from the traditional and one-step PCR methods.

HR represents homologouse recombination.

4.3.3 Improving HR:NHR ratio of transformants obtained by one-step PCR method.

The successful application of targeted gene disruption and replacement is dependent on the ratio of homologous to illegitimate recombination (NHR) events during integrative transformation (Zorin et al., 2005). This section describes experiments aimed at improving the ratio of homologous to non-homologous recombination using PCR products from a one-step PCR method.

4.3.3.1 Using ssDNA for transformation

Several reports in literature have demonstrated the improvement of the HR/NHR ratio by using single stranded DNA for transformation. Such improvements have been shown in the single celled eukaryote microorganisms; *Chlamydomonas* (Zorin et al., 2005) and *Saccharomyces cerevisiae* (Simon and Moore, 1987) and in the single celled cyanobacterium *Cyanothece* sp. strain PCC 7822 (Min and Sherman, 2010a). Work done in these organisms provided some evidence that single stranded DNA (ssDNA) is more efficient than double stranded DNA (dsDNA) for targeted recombination.

It is hypothesised that the use of ssDNA might prevent illegitimate integration processes (Min and Sherman, 2010b). For example, when using ssDNA to transform *C. reinhardtii*, non-homologous DNA integration appeared to be more than 100-fold reduced compared with the use of dsDNA, thus allowing the isolation of homologous recombinants in this organism (Zorin et al., 2005).

During transformation in *Synechocystis*, a Ca^{2+} dependent nuclease located in or on the cytoplasmic membrane degrades one of the two strands of DNA during uptake resulting in the incorporation of ssDNA (Barten and Lill, 1995). Although *Synechocystis* can be transformed with either single- or double-stranded DNA (Vermaas, 1998), DNA taken up by the cell is always in the form of ssDNA (Kufryk et al., 2002). In the following section, ssDNA was transformed into *Synechocystis* to determine whether HR/NHR ratios could be improved. The *psbN* knockout PCR product with 1000 flanks, generated in section 4.3.1.1 was used as a template to generate either ssDNA or dsDNA. Template PCR product (300 ng) was used for the amplification of ssDNA using only one primer - either the 1000 flanks forward primer or the 1000 flanks reverse primer, while 3 ng of the PCR product was used for the amplification of the dsDNA using both and

primers. The same positive and negative controls were included in the PCR reactions as those used in section 4.3.2.3 for band size comparison.

The PCR products were purified and the ssDNA were treated with *Hind*III to get rid of possible trace contamination of the template dsDNA. Transformation was done according to the standard protocol and transformants were selected on 50 µg/ml kanamycin-containing medium. The transformation resulted in many colonies, from which four of each transformation type were selected for PCR analysis to check for targeted *psbN*::km^R knockout as described in section 4.3.1.3. PCR results presented in Figure 4.14 show that all transformants analysed were targeted *psbN* gene knockout (3.1 kb). This confirms the feasibility of generating targeted gene knockouts in *Synechocystis* using ssDNA.

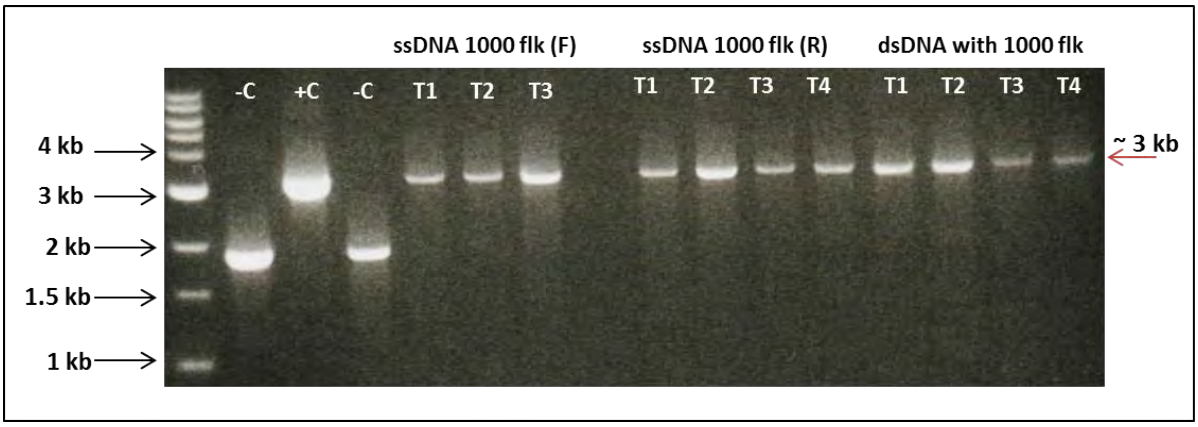


Figure 4.14: PCR results of *Synechosystis* transformed with single stranded DNA.
Transformants generated using ssDNA of the sense primer (F), none sense primer (R) and dsDNA all with 1000 flanks homology to *psbN*. All transformants are *psbN*::km^R knockouts (3.1 kb).This result indicates the possibility of using ssDNA for targeted gene knockouts in *Synechosystis*. The controls in order are: plasmid A without a km cassette (-c), plasmid A (+c) and genomic 6803 DNA (-c).

The next step involved using ssDNA and dsDNA with shorter flanks (100 bp or less). The same protocols were conducted for the generation, purification and transformation of the ssDNA and dsDNA PCR products with short flanks. Transformants were obtained for both ssDNA and dsDNA with 100 flanks. Interestingly, a single transformant was obtained with ssDNA with 50 bp flanks using forward and the reverse primers, while no transformants were obtained with the dsDNA of the same size flanking regions. Finally, using ssDNA with 30 bp homology resulted in no transformants, while four colonies were obtained with

dsDNA of the same size. The variability in obtaining transformants using both the ssDNA and dsDNA when using flanking regions less than 100 bp is similar to the results obtained previously in Figure 4.5 that showed inconsistency of the colony counts with PCR products with regions less than 100 bp.

PCR analysis of the transformants for *psbN* knockout was done as described in section 4.3.1.3 and revealed that the majority of the transformants were not *psbN* knockouts. One exception was of the ssDNA with 50 bp flanks obtained using the forward primer (Figure 4.15). This experiment showed that using ssDNA with short flanks (<100 bp) did not show any improvement to the HR:NHR. Furthermore, the fact that even the dsDNA obtained in this traditional method did not result in targeted *psbN* knockout further supports the hypothesis put forward in Figure 4.13.

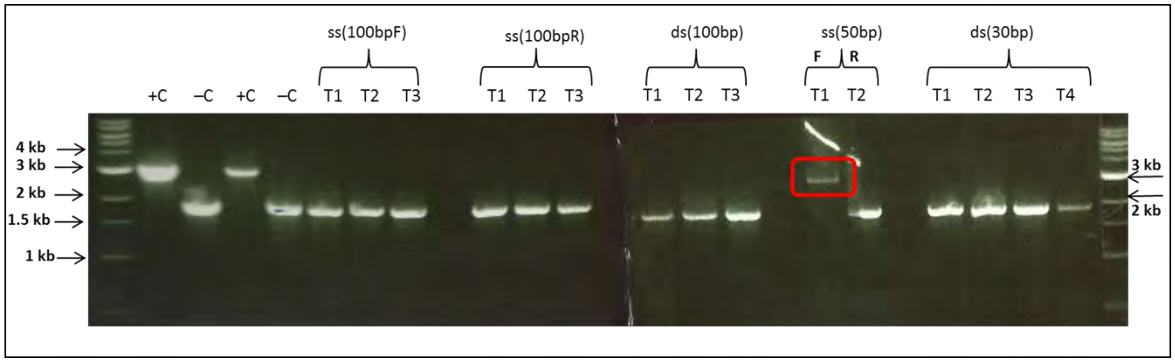


Figure 4.15: PCR results of *Synechocystis* transformed with single stranded DNA with short flanking sequences.

Only one of the ssDNA transformations resulted in a *psbN*::km^R knockout (ssDNA with 50 bp flanks T1). In this case even dsDNA did not result in a positive knockout. The controls in order are: plasmid A (+c), plasmid A without a km^R cassette (-c), *psbN*::km^R knockout (+c) and genomic 6803 DNA (-c).

4.3.3.2 Varying DNA and cell concentration used in transformation

Typically for *Synechocystis* transformation, 1-5 µg of DNA is added to recipient strains at a concentration of 4 x 10^8 cells/ml (Zang et al., 2007, Kufryk et al., 2002, Lindberg et al., 2010). In this section, the effect of varying DNA concentration and recipient cell concentration on the ratio of HR:NHR was investigated.

For this experiment, LT30 PCR products were generated using standard PCR conditions described in section 4.3.2.1. It took many PCR attempts in order to amplify 9 µg of the LT30.*psbN*::km^R knockouts using the one-step PCR method.

This amount was split in the transformation to be used as 0.5 µg, 1 µg, 1.5 µg and 2 µg of PCR product added to 1 ml of WT recipient cells containing 4 x 10^8 cells. On the other hand, 1 µg of the same DNA was added to 1 ml of WT recipient cells containing 100%, 50%, 25% and 12.5% of WT recipient cells (with 100% equal to a cell count of 4X10^8 cells). This experimental set up is illustrated in Figure 4.16.A. Furthermore, 1 ml of a negative control (WT recipient strain) was included in the experiment that contained 4X10^8 cells with no added DNA. This was used to test whether colonies appear due to spontaneous resistance to the antibiotic.

Transformation was carried out according to methods described in section 2.4.2. The transformants were allowed to recover for 48 hours before overlaying them with BG11 containing 50 µg/ml kanamycin. Of the hundreds of colonies visible on the plate, seven to ten days post antibiotic selection, 37 were re-streaked on kanamycin selective medium out of which only ten survived (Figure 4.16.A). PCR analysis done on the surviving colonies, using the same PCR conditions and primers described in section 4.3.1.3, showed that only one of the ten colonies screened was a *psbN* knockout (Figure 4.16.B). The results show that varying the concentration of DNA and recipient cell concentration does not seem to have an observable effect on the HR:NHR, at least at the DNA and cell concentrations used. Nevertheless, a wider range of DNA concentrations should be used and more colonies screened in order to gain a more conclusive result.

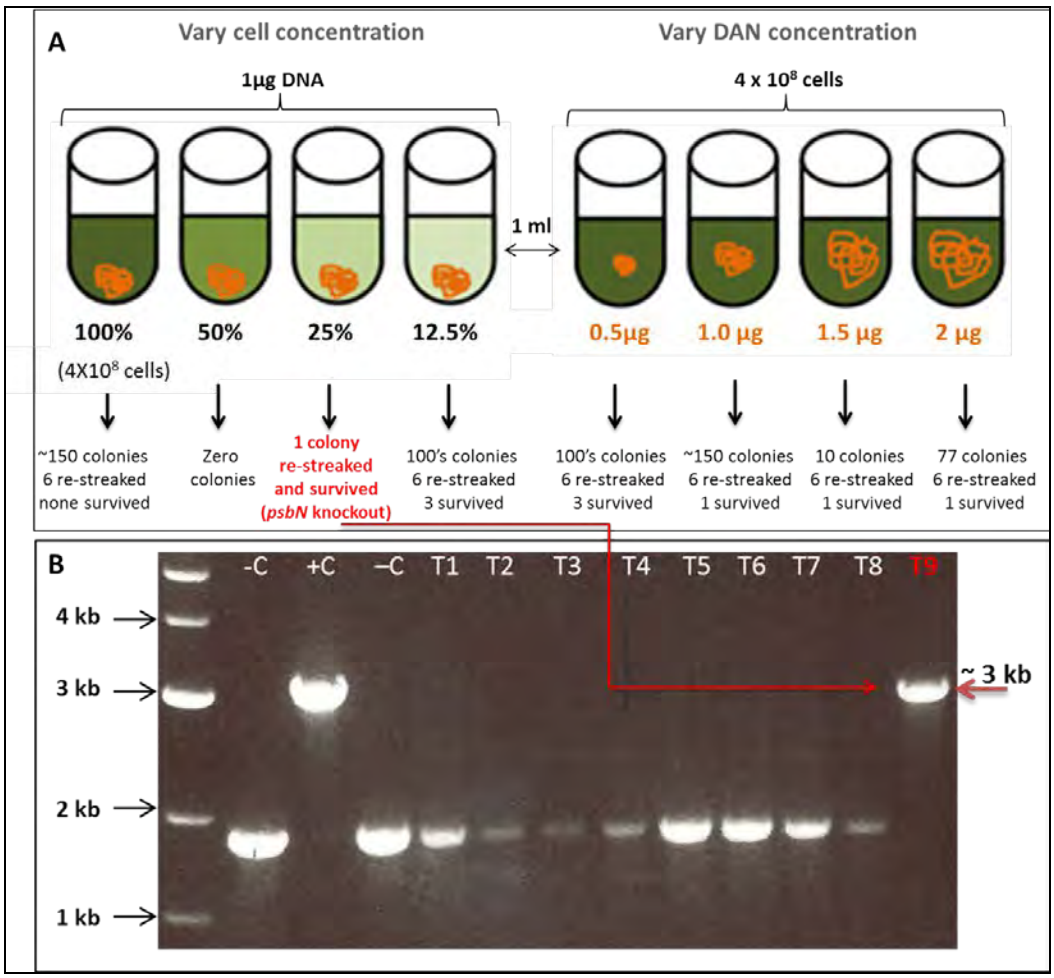


Figure 4.16: Transformants & PCR results of transformants obtained by varying DNA and cell concentration.

(A) Schematic of *Synechocystis* transformation using varying DNA and cell concentrations, as described in section 4.3.3.2. (B) PCR results show that a single *psbN*::km^R knockout was obtained, indicating that changing DNA concentration or concentration of cells used in the transformation had no observed effect on improving HR:NHR. The controls in order are: genomic 6803 DNA (-c), plasmid A (+c), plasmid A without a km cassette (-c), T1-T9 represents the transformants. [Note: T10 which was not a *psbN* knockout is not included in this figure].

4.3.3.3 Using Δ*recJ.*ble as recipient strain for transformation.

As discussed in chapter 3, the RecJ-deficient strain displayed higher transformation efficiency due to the deletion of the *recJ* gene that encodes an exonuclease responsible for foreign DNA degradation. In this work, this "super strain" was used as a recipient strain as opposed to WT cells to determine whether this would improve the HR:NHR ratio. Furthermore, since *Synechocystis* can resist high concentrations of kanamycin (up to 200 µg/ml) as seen in earlier spot tests (Figure 4.12.B); selection of transformants was carried out using 200 µg/ml of the antibiotic. This was done to reduce background colonies that may

arise due to spontaneous resistance. Finally, PCR products of *psbN*::km^R knockout with 50 bp flanks were used as opposed to 30 bp in previous work to improve the integration process.

PCR was carried out to amplify the *psbN*::km^R PCR products with 50 bp flanks (LT50) from the pUC4K plasmid as in section 4.3.2.1 using a 'hot start' in an attempt to generate the PCR product more easily. The PCR product was purified and treated with *Dpn*I and 7 µg used in the transformation of both WT and ∆*recJ*.ble recipient strains. WT and Δ*recJ*.ble strains were also transformed with a PCR product generated from the pUC.psbN.km^R plasmid with 1000 bp of flanking sequence (as a positive control), and with no DNA (negative control). A large number of putative transformants (116 kanamycin resistant colonies) appeared within seven days of antibiotic selection using the RecJ- recipient strain, while no transformants were obtained from the WT recipient strain, confirming the higher transformation efficiency of the RecJ- strain.

Sixty isolated colonies were selected for analysis after undergoing three rounds of selection on medium supplemented with 200 µg/ml of km. DNA extraction from the colonies was done in such a way that every ten colonies were pooled in a single preparation and used for PCR analysis. PCR analysis on the pooled DNA extracts was carried out to check for targeted *psbN* gene knockout as described in section 4.3.1.3. The PCR results confirmed the presence of *psbN* knockouts in each of the pooled DNA preparations (Figure 4.17.A). To investigate the proportion of *psbN* knockouts in the pooled samples, twelve representative colonies (two from each pool) were randomly selected and individual genomic preparations were made for PCR analysis. Results in Figure 4.17.B show that all twelve individual colonies screened were positive for the *psbN*::km^R knockout.

After confirming the ability to generate a large number of *psbN* gene knockouts by means of homologous recombination using the one-step PCR method with an LT50 product and using the *ΔrecJ*.ble as the recipient strain, it was necessary to check for reproducibility of these results. Therefore, the same experiment was repeated three times; however, results obtained were not consistent with our initial findings as very few colonies were isolated and most were not *psbN*::km^R knockouts (Table 4.1).

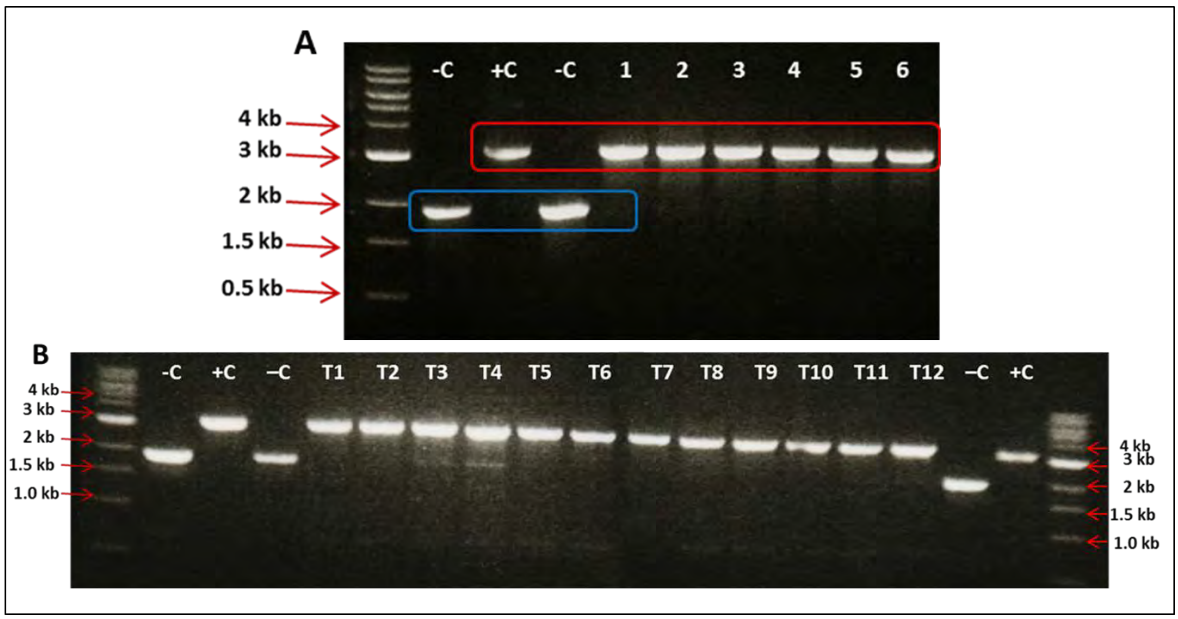


Figure 4.17: PCR analysis of transformants in Δ*recJ*.km recipient strains.

(A) The transformation of Δ*recJ*.km strain with PCR products of *psbN*::km^R with 50 flanks resulted in *psbN*::km^R knockouts within pooled DNA of 60 tested transformants. Each lane (numbered 1-6) is a PCR result on DNA pooled from 10 transformants. The controls in order are: genomic 6803 DNA (-c), plasmid (+c), plasmid without a km^R cassette (-c). (B) 12 random transformants (T1-T12) from those analysed in (A) were selected to test for gene knockouts in the Δ*recJ*.km strain and all showed a positive result for the *psbN*::km^R knockout.

Table 4.1: Transformation results of LT50.*psbN*::km^R PCR products in Δ*recJ*.ble and WT recipient cells.

TRANSFORMATION #	RECIPIENT STRAIN	# OF TRANSFORMANTS	*PSBN* KNOCKOUT
1	Δ*recJ*.ble	116	60 pooled DNA
	WT	0	-
2	Δ*recJ*.ble	1	YES
	WT	0	-
3	Δ*recJ*.ble	2	NO
	WT	0	-
4	Δ*recJ*.ble	3	NO
	WT	0	-

The failure to reproduce successful *psbN*::km^R knockouts raised a question as to whether there is something about the *psbN* locus itself that might be playing a role in preventing consistent HR. Therefore, two other genes, *psbA2* (gene code=slr1311) and *psb28.2* (gene code=slr1739) were selected for targeted gene knockouts in the Δ*recJ*.km^R strain. The two genes were selected on the basis

that, like *psbN*, they are both benign and deleting them should not result in any phenotypic change or effect on the physiology and growth of the deleted strain.

Again for this PCR, primers were designed such that 22 bp were homologous to the kanamycin resistance cassette at their 3' end and ~50 bp were homologous to *psbA2* and *psb28.2* genes at their 5' ends. Primers used for this PCR can be found in Appendix 3.

The PCR was performed as before (section 4.3.2.1) using pUC4K as the template for the PCR reaction. For both sets of primers the km^R gene was successfully amplified with ~50 bp flanks on either end, generating the expected ~1.3 kb gene product. A *psbN*::km^R product with 50 bp flanks was also produced, together with the *psbA2*::km^R and *psb28-2*::km^R products. The products were purified and treated with *Dpn*I, after which 3-8 µg of DNA was used for transformation into both WT and Δ*recJ*.ble recipient strains. Negative and positive controls for both strains were included in the transformation for band size comparisons. The details of the controls used are in the legend of Figures 4.18 and 4.19.

Putative colonies appeared seven to ten days post antibiotic selection (200 µg/ml of km) from all three gene transformations using the *ΔrecJ*.ble recipient strain, whereas no colonies were observed for the WT recipient strain. Approximately 150 colonies were obtained from the LT50.*psbA2*.km^R PCR product, five colonies from LT50.*psbN*::km^R and ~60 colonies from LT50.*psb28.2*.km^R. For PCR analysis of the putative transformants for targeted gene knockouts, sixty isolated colonies were selected from both LT50.*psbA2*.km^R and LT50.*psb28.2*.km^R where every ten colonies were pooled in single genomic preparations. Genomic DNAs from the five colonies from LT50.*psbN*::km^R transformants were prepared individually. Primers were designed upstream and downstream of each tested gene to check for integration of the PCR products in the *psbA2* and *psb28.2* loci. The expected band size of *psbA2* knockouts was ~1.4 kb (722 bp for WT) and ~1.5 kb for *psb28.2* knockouts (755 bp for WT), and as before 3.1 kb for *psbN* knockouts (1.9 kb for WT). Primers used in this PCR analysis are listed in Appendix 3.

PCR analysis of the transformants was carried out according to methods described in section 4.3.2.1. Results presented in Figure 4.18 showed that none of the 60 colonies screened for targeted gene knockouts for the two genes *psbA2*

or *psb28.2* were positive, while only one of the five transformants screened for *psbN* gene knockout was positive.

Next, four random colonies were selected from each of the LT50.*psbA2*.km^R and LT50.*psb28.2*.km^R transformants for individual PCR analysis to confirm the absence of genes knockouts. Again, none of the individually screened transformants showed targeted gene knockouts. From these results we can conclude that, at least for the three genes selected in this study, the gene locus did not appear to affect the integration pattern of the PCR products with short flanks. Further analysis was carried out to find out where the PCR products were being inserted when not integrated via double homologous recombination. It was hypothesised that the PCR products may be integrating in the same “hot spot” identified in chapter 3 (*speA*/*psbA2* locus). It is possible that the short flanks of the PCR products sometimes fail to locate their homologues in the genome, therefore acting like the naked km^R cassette landing in these "hot spots". To test this theory, a PCR was carried out where DNA from pooled genomic preparations of the LT50.*psbA2*.km^R and LT50.*psb28.2*.km^R transformants and the five individual DNA preparations from the LT50.*psbN*::km^R transformants were used as templates. Primers were designed where one primer was located upstream of the hot spot area (*speA*/*psbA2* area) and another located within the km^R cassette. Obtaining a PCR product using this set of primers would confirm the integration of the PCR product in the *speA/psbA2* locus.

The PCR was carried out using primers listed in Appendix 3. Wild-type genomic DNA was used a negative control, while genomic preparations of the Δ*recJ*.km.*Hinc*II strain (obtained from work carried out in chapter 3) was used as a positive control since the km^R cassette in this transformant was found to be located in the *speA*/*psbA2* area. Results presented in Figure 4.19 show that some of the PCR products did indeed integrate in the *speA*/*psbA2* area. Three of the five LT50.*psbN*::km^R PCR products and one of the sixty LT50.*psb28.2*.km^R PCR products were located in the *speA/psbA2* region, while none of the LT50.*psbA2*.km^R PCR products were found in that area and have probably integrated elsewhere in the genome. We can conclude from the previous results that when non-homologous recombination events take place in *Synechocystis* using PCR products of gene knockouts with short flanks, some PCR products act

as naked km^R cassettes and integrate into hot spots in the genome, while others integrate randomly in other locations.

After testing several parameters for the improvement of the HR:NHR ratio (*e.g.* using ssDNA, varying DNA and recipient cell concentrations and using a modified recipient strain of *Synechocystis*) and combining the results with those obtained in chapter 3 (similar integration patterns between PCR products of km^R gene knockouts with short flanks to those of a naked km^R cassette), it was concluded that the sequence of the km^R cassette itself may favored integration into random or hot spots in the genome, similar to results obtained in Chapter 3. However, before exploring this hypothesis, it was necessary to compare the two kinds of PCR products that are, at least in principle, identical but give different transformant outcomes in terms of HR;NHR ratios. PCR products obtained from a knockout plasmid (pUC.psbN.Km^R) and from the one-step method (with 5' extensions homologous to *psbN*), both with 50 bp flanking regions, were transformed into *Synechocystis* for direct comparison of different types of PCR products (as illustrated in Figure 4.20).

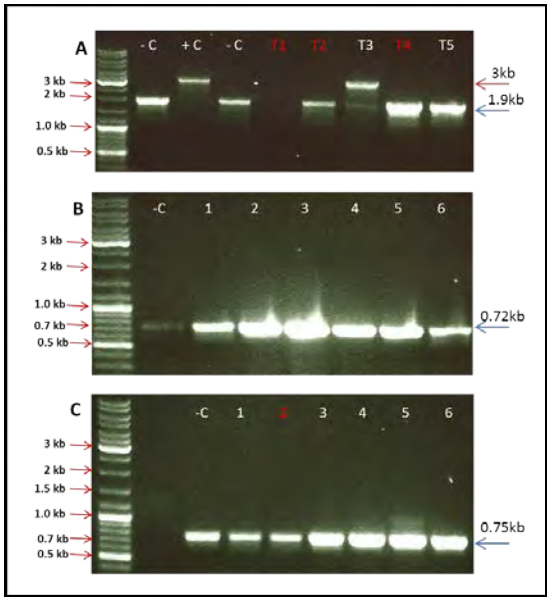


Figure 4.18: Applying the one-step PCR to knock out three genes using Δ*recJ*.km as recipient strain.

(A) *psbN*::km^R knockout (3.1 kb). The controls in order are: (-C): plasmid carrying WT *psbN* gene (plasmid A without km^R), (+C): pUC.psbN::km^R, (-C): genomic 6803 DNA. (B) PCR results show no positive *psbA2*.km^R knockouts (WT band size ~ 0.72 kb) (-C): genomic 6803 DNA. (C) PCR results show no positive *psb28*.2.km knockouts (WT band size ~ 0.75kb). (-C): genomic 6803 DNA. Each lane (numbered 1-6) is a PCR result on DNA pooled from 10 transformants (for both B and C).

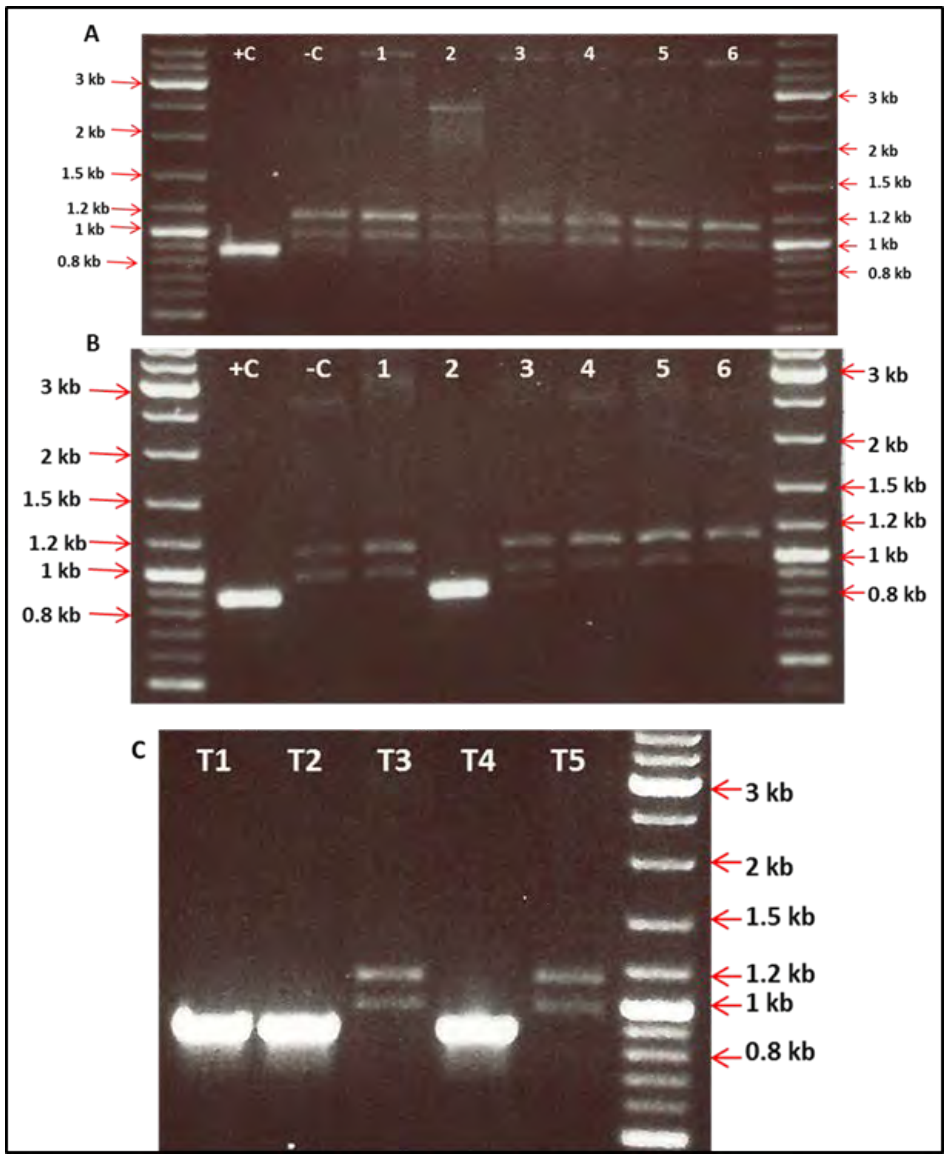


Figure 4.19: Testing the integration of PCR products of three genes with short flanks in the *speA/psbA2* area.

(A) None of the LT50.*psbA2*.km^R transformants are in the *speA.psbA2* area. (B) Only one of the LT50.psb28.2.km transformants was integrated in the *speA.psbA2* area. Individual analysis of the pooled DNA in preparation # 2 showed only 1/10 pooled DNA integrated in *speA.psbA2* area. (C) Three transformants of LT50.*psbN*::km^R (1, 2 & 4) integrated in *speA.psbA2* area. Each preparation (numbered 1-6) is a PCR result on DNA pooled from 10 transformants. +C: Δ*recJ*.km.*Hinc*II strain that contains a km^R cassette in the *speA.psbA2* area, (-C): genomic 6803 DNA.

4.3.4 Comparison of PCR products from the traditional and one-step PCR methods:

In order to understand the reason behind the different modes of integration of what ought to be the same final PCR product derived from both the traditional

method (section 4.3.1.1) and one-step PCR method (section 4.3.2.1), it was necessary to compare the steps undertaken in generating the PCR products from each method. Figure 4.20 illustrates the PCR product of a *psbN*::km^R knockout produced via the traditional method and the one-step PCR method. From the figure, we can conclude that there are two apparent differences in obtaining the final product that may play a part in the difference in integration patterns between the two PCR products; namely, the template used for amplifying the gene knockout and the primers used for the amplification.

With regard to the template used in the traditional method, one may argue that there is a possibility of trace amounts of the original pUC.psbN::km^R plasmid construct that may be contaminating the PCR product and giving rise to the transformant colonies via double homologous recombination. However PCR products generated from the plasmid were treated with *Dpn*I that recognises the sequence GATC and digests it when methylated [=plasmid DNA], but doesn't digest the un-methylated sequence [=PCR products], as illustrated in Figure 4.12. Furthermore, the possibility that transformants were obtained from trace amounts of plasmid that might have escaped the *Dpn*I digestion can be ruled out since the number of colonies obtained from transformation with each PCR product (with varying flank sizes) was proportional to the length of regions of homology to the target gene carried on the PCR product suggesting that transformation arose from the PCR products and not from traces of the plasmid construct (Figure 4.5).

Next, the primers used in the one-step PCR method were examined. It is known that PCR involving long-tail primers are prone to be more complex since longer primers are expected to be contaminated with incomplete primers that are not full-length (Datsenko and Wanner, 2000). Therefore, to investigate the effect of employing long-tail primers in amplifying gene knockouts and to check for the addition of proper flanks to the km^R cassette in the generated PCR products, the product of *psbN*::km^R with 50 bp flanks obtained from the one-step PCR and that obtained by amplification from pUC.psbN.Km^R were sequenced and compared. Interestingly, sequence results revealed that the junction between the km^R cassette and the flanking sequences (homologous to *psbN*) are not the same in the two PCR products and nucleotide deletions and additions are found at the junction in PCR products obtained from the one-step PCR method as shown in Figure 4.22.A.

Interestingly, a search of the BLAST database with a sequence of 30 bp present in the junction between the part of the long-tail primer that primes to the km^R cassette (highlighted in grey in Figure 4.22.B) and the part of the long-tail primer representing the *psbN* flank (highlighted in yellow in the same figure) is part of the km^R cassette itself. It is unclear why this recombination/shuffling took place at the junction but it seems to be due to issues related to the sequence of the km^R cassette. From the sequence results obtained, it seems reasonable to conclude that the PCR products generated using long-tail primers (one-step PCR method) combine less efficiently due to the deletions and mutations that take place during the amplification of the PCR product.

To further investigate the effect of using long tail primers, the PCR product of LT50.*psbA2*.km^R was also sequenced to examine the integration of the correct flanks in this PCR product. As seen with the LT50.*psbN*.km^R PCR product, the sequence results confirmed the presence of nucleotide deletions and mutations in the PCR products within the long-tail primers (Appendix 7). A study done by Datsenko and Wanner. (2000), showed that a single nucleotide deletion that occurred in long-tail primers (≥60 bp) happened very near the junction of the priming site and homology extension which resulted in less efficient recombination of the PCR products.

In our case, four nucleotide deletions took place at the junction and therefore it is arguable that the recombination of the PCR product is even less efficient. These kinds of errors are difficult to avoid since oligonucleotides with internal deletions arise from chemical synthesis and are difficult to remove by conventional purification methods (Temsamani et al., 1995). Alternatively, the problem might arise from secondary structure formation in the primers during PCR amplification. The mis-matches/deletions in the PCR product could explain why targeted gene knockouts were not obtained using the one-step PCR method while targeted *psbN*::km^R knockouts were obtained with PCR products amplified from the plasmid construct. On the other hand, it could also be possible that the inefficient recombination (NHR) of the one-step PCR product was due to the DNA shuffling that took place between the km^R cassette and the long-tail primer (Figure 4.22.B) and that there were high chances of generating errors in the long-tail primers when the km^R is used. The following section explores the possibility of using

another marker instead of the km^R cassette to circumvent these apparent problems.

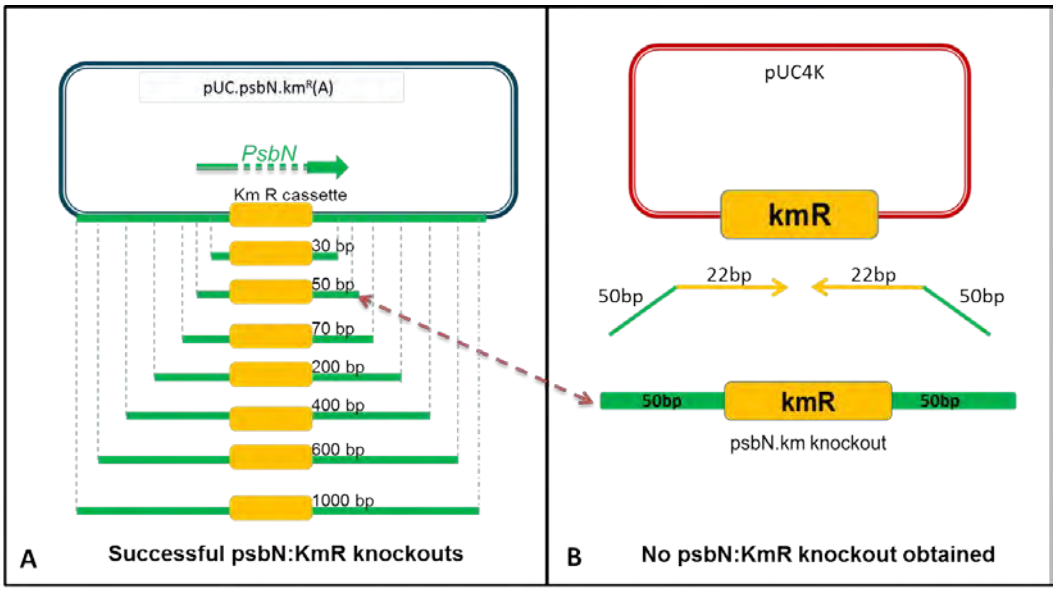


Figure 4.20: Illustrations of PCR products obtained from traditional and one-step PCR methods.

(A) PCR products amplified from plasmid construct containing *psbN*::km^R knockout using the traditional method (B) PCR product of *psbN*::km^R knockout amplified directly from plasmid carrying the kanamycin resistance cassette using the one-step PCR method.

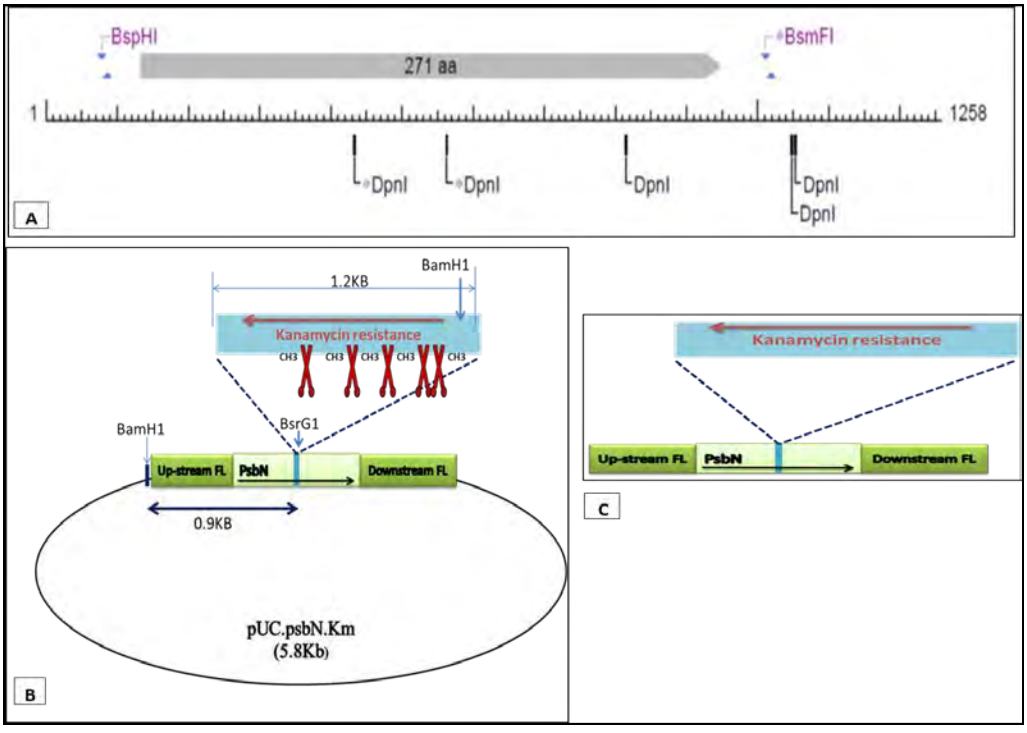


Figure 4.21: Illustration of *DpnI* digestion of km^R cassette in pUC.psbN::km^R plasmid.

(A) Km^R cassette showing *Dpn*I sites. (B) Plasmid digested with *Dpn*I at methylated sites. (C) PCR product protected from *Dpn*I digestion (Nonmethylated).

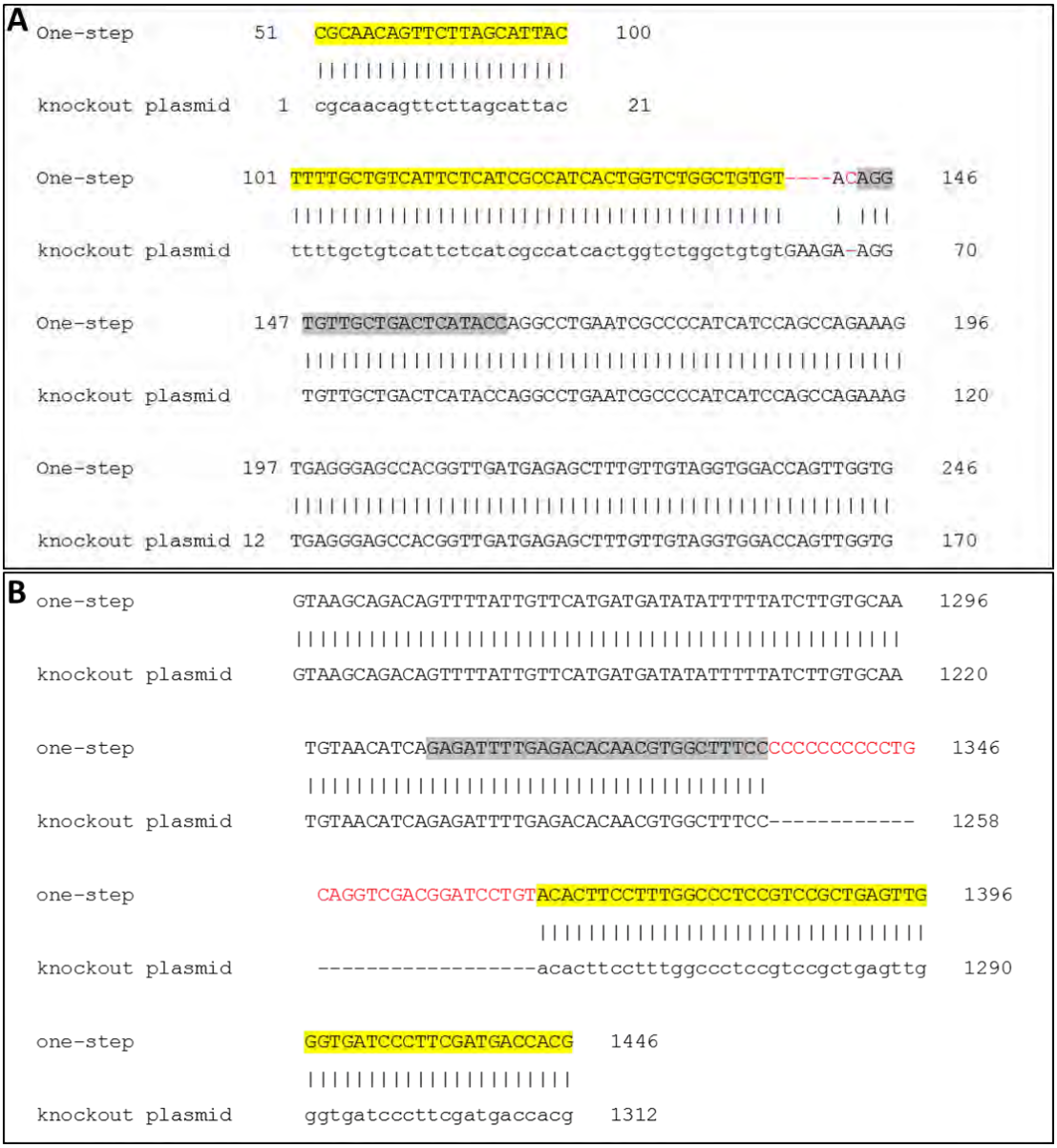


Figure 4.22: Sequence results of PCR products obtained by one-step PCR and traditional methods.

Sequence analysis shows that the junction area between the long-tail primer segment homologous to the target gene (highlighted in yellow) and that homologous to the selectable marker (grey) are not the same in PCR products obtained from the knockout plasmid and the one-step PCR method. (A) The junction shown in red font contains four nucleotide deletions and an extra nucleotide. (B) Blast results of the 30 bp extra nucleotides highlighted in red were found to be part of the km^R cassette that has been inserted in the junction between the part of the long-tail primer that primes with the km^R cassette (light grey) and the *psbN* flank (yellow).

4.3.5 Utilisation of Δ*recJ*.km^R strain and new gene knockouts for one-step PCR

In order to test whether the choice of the selectable marker (km^R cassette) was responsible for the NHR events taking place in *Synechocystis*, the *ble* selectable marker (encoding zeocin resistance) was chosen to replace the km^R cassette in creating targeted *psbA2*, *psb28.2* and *psbN* gene knockouts using the one-step PCR method. The following section describes the utilisation of the Δ*recJ*.km^R mutant strain (developed in chapter 3, section 3.3.4) and the *ble* marker in creating the three gene knockouts.

4.3.5.1 One-step PCR gene deletion using *ble* as the selectable marker.

Long-tail primers were designed to amplify the *ble* resistance cassette directly from plasmid pZΔES (Appendix 1) as *psbN*.ble, *psbA*2.ble and *psb28.2*.ble gene knockouts. The long-tail primers were designed such that ~22 bp were homologous to the *ble* cassette at their 3' end and ~50 bp flanks were homologous to the target gene (*psbN*, *psbA2* and *psb28*.2) at their 5' ends. PCR was performed as described in section 2.3.5 and results presented in Figure 4.23 show the successful amplification of all three LT50 products in a single PCR step. The purified PCR product (0.6-0.7 kb) was *Dpn*I-digested and used for the transformation into the Δ*recJ*.km^R strain of *Synechocystis* as described in section 4.3.4.1. Sequences of the long-tail primers used can be found in Appendix 3.

4.3.5.2 Transformation of Δ*recJ*.km^R strain with the one-step PCR products

The PCR products obtained in section 4.3.5.1 were used for transformation into both WT *Synechocystis* and the Δ*recJ*.km^R mutant as in section 2.4.2. Transformants were allowed to recover for 48 hours before selection with zeocin at a final concentration of 7.5 µg/ml. Approximately 60 colonies were recovered (7-10 days post antibiotic selection) upon transformation of the Δ*recJ*.km^R strain with the PCR products, while no colonies appeared on the negative control, indicating that the obtained colonies are most likely true transformants.

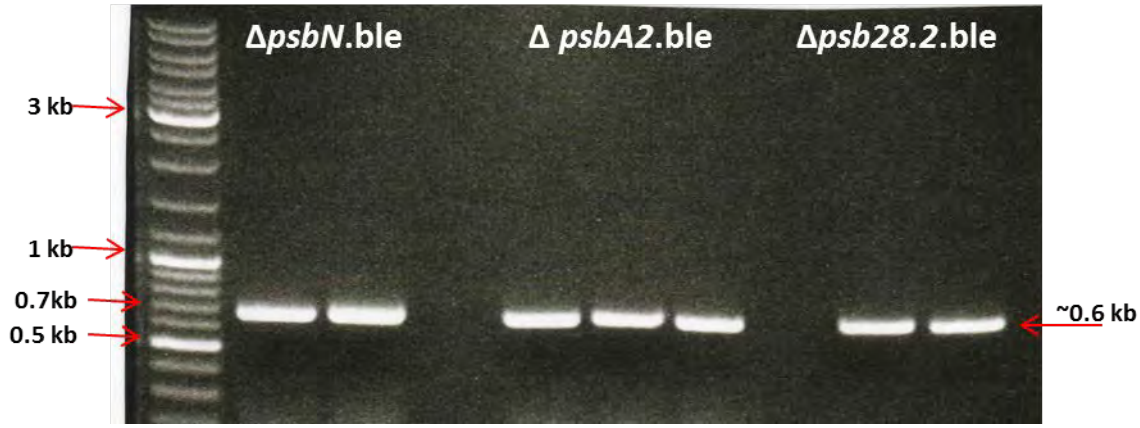


Figure 4.23: PCR products of *psbN*, *psb28.2* and *psbA2* gene knock using *ble* marker gene.

PCR products of *psbN*, *psb28.2* and *psbA2* gene knockouts using long tail primers and ble as the selectable marker (one-step PCR products). Expected product sizes = ~0.7 kb.

Interestingly, the number of transformants varied between the three transformations with more transformants obtained with PCR products of LT50.*psbN*.ble PCR product (48 colonies) followed by LT50.*psbA2*.ble PCR product (six colonies) while only two colonies were obtained using LT50.*psb28.2*.ble PCR product (Figure 4.24). The variability in the colony count is probably due to differences in the sequences of each gene that may affect integration. On the other hand, transformation of the WT strain yielded only a few transformants for LT50.*psbN*.ble, LT50.*psbA2*.ble and LT50.*psb28.2*.ble PCR products (10, 15 and a single colony, respectively). Additionally, many colonies appeared on the negative control plate, indicating that the colonies generated from the WT strains are probably false transformants that appeared due to spontaneous resistance or local depletion of the antibiotic on the plate.

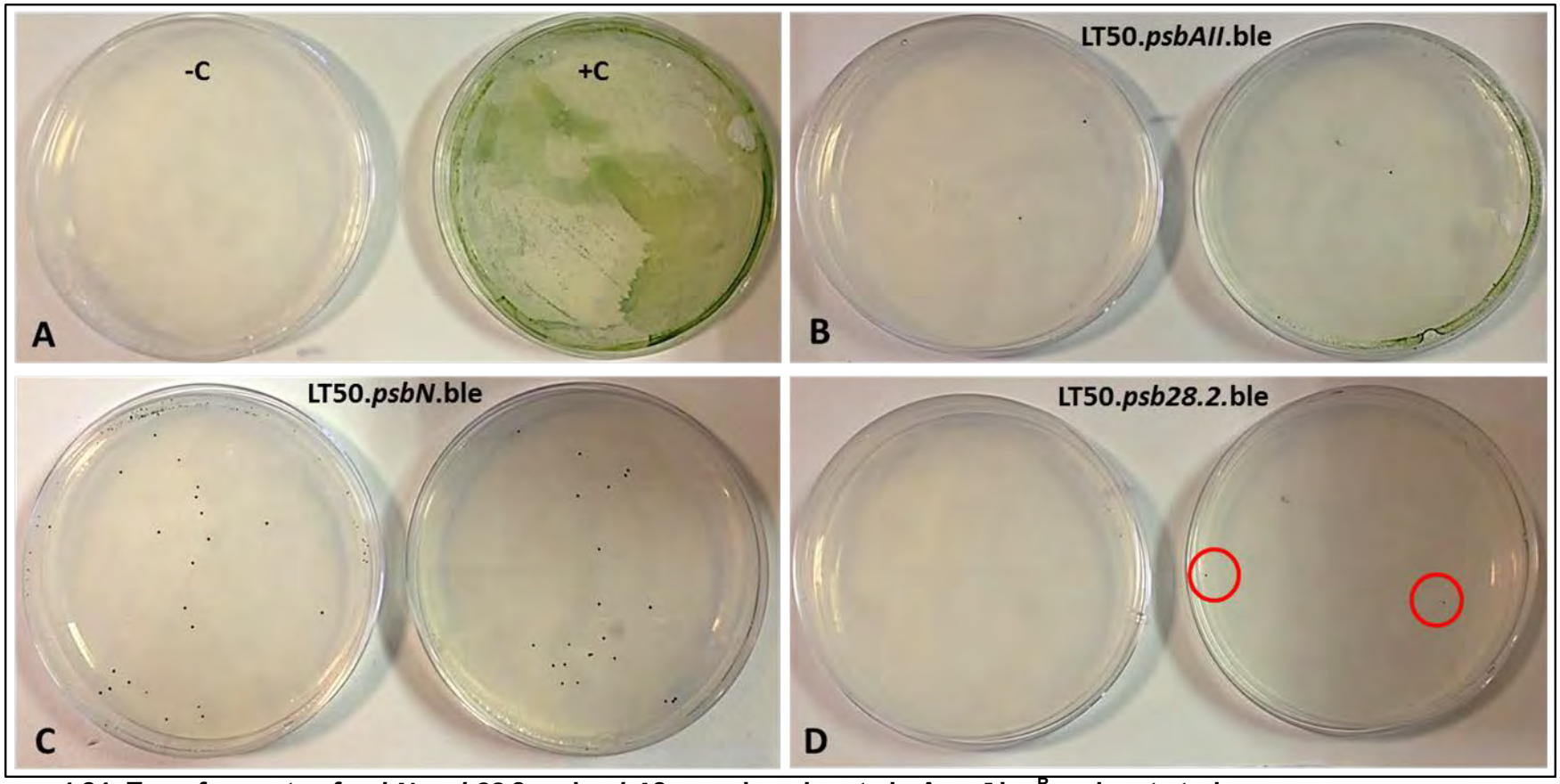


Figure 4.24: Transformants of *psbN*, *psb28*.2 and *psbA*2 gene knock-outs in Δ*recJ*.km^R recipent strain.

(A) Negative and positive controls (strain with no added DNA, strain transformed with Δ*psb28.2*.ble respectively) of the Δ*recJ*.km strain showed expected results. (B) Six colonies obtained from transformation of Δ*recJ*.km^R strain with *psbA*2.ble PCR product. (C) 48 colonies obtained from transformation of Δ*recJ*.km^R strain with *psbN*.ble PCR product. (D) Two colonies obtained from transformation of Δ*recJ*.km^R strain with *psb28.2*.ble PCR product.

4.3.5.3 Analysis of transformants for the three gene knockouts in Δ*recJ*.km^R strain.

Genomic DNA of selected transformants obtained in section 4.3.5.2 was prepared and used as a template in a PCR reaction using primers designed to check for the correct integration of the PCR products into the targeted gene loci (one primer upstream of the gene of interest and other designed within the *ble* marker gene). Results presented in Figure 4.25 shows the successful isolation of gene knockouts for all three genes tested. As seen in the figure, 5 out of 6 tested transformants were *psbA2.*ble knockouts, 6 out of 10 tested transformants were *psbN.*ble knockouts and the single colony obtained with LT50.*psbA2.*ble PCR product was also a knockout. On the other hand, none of the transformants obtained using the WT recipient strain were positive gene knockouts for any of the three genes tested, confirming that they were false transformants. Positive (for *psb28.2*) and negative controls (WT genomic DNA) were included in some of the PCR reactions. Primers used to test for the transformants can be found in Appendix 3.

From the results obtained above, we can conclude that the one-step PCR method can be efficiently used to create targeted gene knockouts in *Synechocystis* when the km^R selectable marker is replaced with the *ble* selectable marker, possibly due to matters related to the sequence of the km^R cassette that favour its recombination into other areas in the genome, as seen in chapter 3. It would be beneficial to be able to use shorter flanks for homology to the target gene, thereby shortening the overall length of the primer and reducing the cost of primer synthesis.

Primers were designed that contained at their 5' ends 30 bases of homology to the flanking regions of *psbN*, *psbA2* and *psb28.2* genes and at their 3' end 22 bases of homology to the *ble* resistance cassette. As before, the *ble* cassette was amplified from pZΔES as a template using the different primer pairs as listed in Appendix 3. The PCR products were used for transformation into both WT and Δ*recJ*.km^R recipient strains. However, no transformants were obtained using 7 µg of the each of the PCR products as shown in Figure 4.26. The transformation was repeated to confirm the findings and the same result was obtained. Our inability to obtain transformants indicated that 50 bases is likely close to the

minimum length of homology needed for this method using *ble* as the selectable marker.

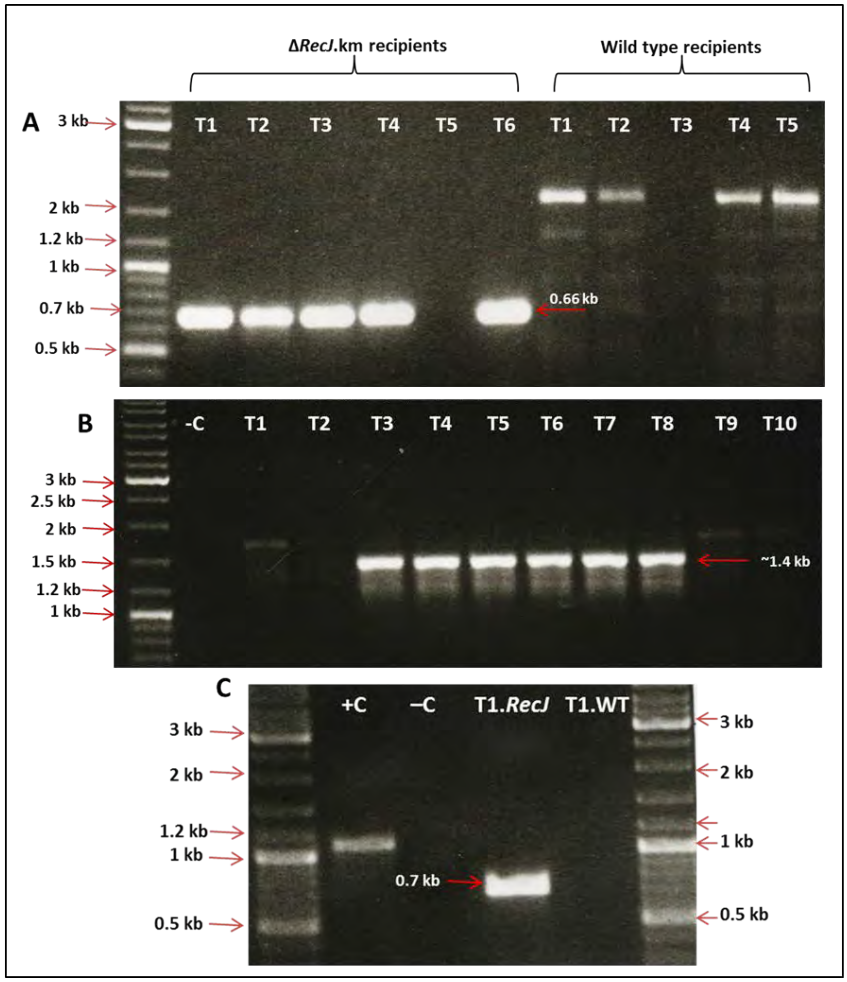


Figure 4.25: PCR analysis of transformants of three gene knockouts in Δ*recJ*.km strain.

(A) 5 out of 6 colonies screened for *psbA2* gene knockout in Δ*recJ*.km^R recipient strains were knockouts while none of the five colonies screened for *psbA2* gene knockout were positive in the WT recipient strain. (B) 6 out of 10 colonies screened for *psbN* gene knockout were positive; (C) The single colony screened for *psb28.2* gene knockout was a knockout while the single colony screened for *psb28.2* gene knockout was negative in WT recipient strain. **Key:** -C; WT genomic DNA, +C; *Synechocystis* with *psb28.2* gene knockouted out with a ble marker.

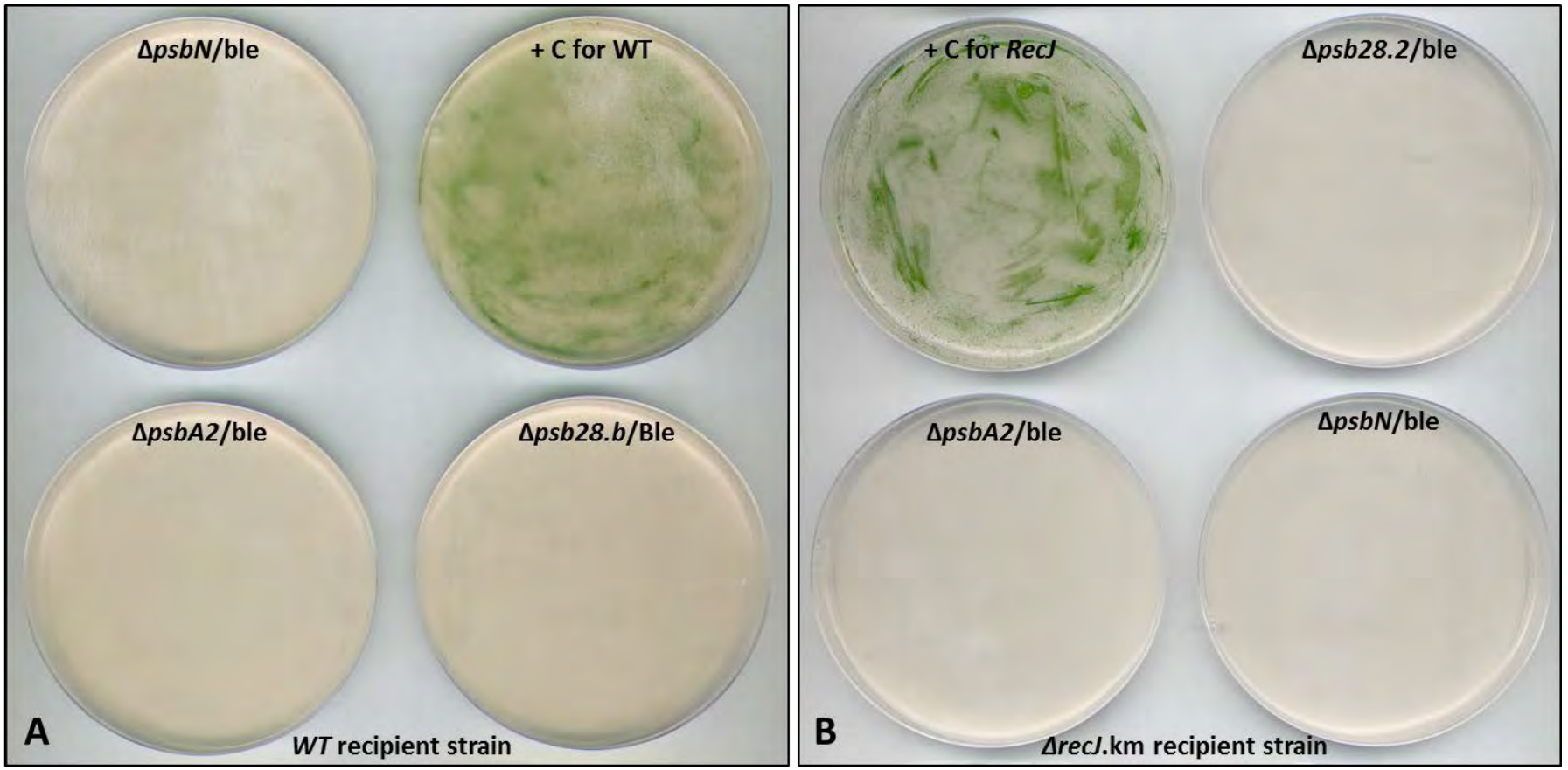


Figure 4.26: Transformation of WT and Δ*recJ*.km^R recipents strains with three gene knockouts with 30 bp flanks from one-step PCR.

(A) No transformants obtained in WT recipient strains from any of the three gene knockouts with 30 bases homology to the target gene. (B) No transformants obtained in Δ*recJ*.km^R recipient strains from any of the three gene knockouts with 30 bases of homology to the target gene. Positive controls show a successful transformation procedure. +C for WT; WT recipient strain transformed with RecJ.ble plasmid, +C for RecJ; RecJ.km recipient strain transformed with plasmid containing a *psb28.2* gene knockout out with a ble marker.

4.3.5.4 Large chromosomal deletions using one-step PCR

After the successful isolation of *psbN*.ble, *psbA2*.ble and *psb28.2*.ble gene knockout mutants using the one-step PCR method, the method was tested for the ability to create large chromosomal deletions (for example, removal of an entire operon). The ability to delete large segments of DNA would benefit from saving immense time in creating knockouts for each gene of an operon. Furthermore, the method can be used to study the effect of operon deletions as well as individual gene deletions.

In this investigation, a four-gene operon consisting of *cpcC1* (gene code: sll1580), *cpcC2* (gene code: sll1579), *cpcA* (gene code: sll1578) and *cpcB* (gene code: sll1577) was selected. The genes of this operon encode components of the phycobilisome complex which is the peripheral light-harvesting complex of cyanobacteria (Vermaas, 1998) and are therefore for required for efficient photosynthesis. It has been reported that the deletion of these four genes would result in a phenotypic change where the colony and culture colour of successful knockouts would be olive green as opposed to the distinct blue-green colour normally observed in *Synechocystis* transformants with benign gene deletions (Lea-Smith et al., 2013). Creating a deletion (whether gene or operon) that would result in such a phenotypic change makes identification of successful transformants a much easier task.

The genome sequence of *Synechocystis* was consulted through Cyanobase to design primers to amplify the target operon. Each primer was designed so that 24 bases were homologous to the *ble* marker at the 3' end and ~50 bases were homologous to the target operon at the 5' end as seen in the schematic representation shown in Figure 4.27. Sequences of these LT50 primers can be found in Appendix 3. PCR was performed as described in section 4.3.4.2 and resulted in the amplification of the correct size of PCR product (~0.7 kb). If successful, the knockout would result in the replacement of the 2 kb chromosomal segment with the ~0.6 kb *ble* selectable marker. The purified PCR product (~4.5 µg) was used for transformation into the Δ*recJ*.km^R strain. Transformants were allowed to recover for 48 hours before being selected on media containing 25 µg/ml zeocin. Only a single colony was obtained 14 days post antibiotic selection. The delay in the appearance of the transformant was

expected since the deletion targeted essential photosynthesis components. The colony was taken through three rounds of antibiotic selection before being analysed by PCR. For PCR analysis, primers were designed upstream and downstream of the targeted deletion (indicated in solid black arrows in Figure 4.27.B) to check for the correct insertion of the one-step PCR product. Primers used to check for the knockouts are listed in Appendix 3. Results of the PCR analysis shown in Figure 4.28 revealed that the tested colony was not a successful knockout indicated by the WT band size of ~2.5 kb (a positive chromosomal knockout would be ~1.2 kb). A second attempt to generate the large chromosomal deletion described earlier was made using the same methods described above. However, no transformants were obtained, indicating that employing the one-step PCR method for large chromosomal deletions may not be feasible.

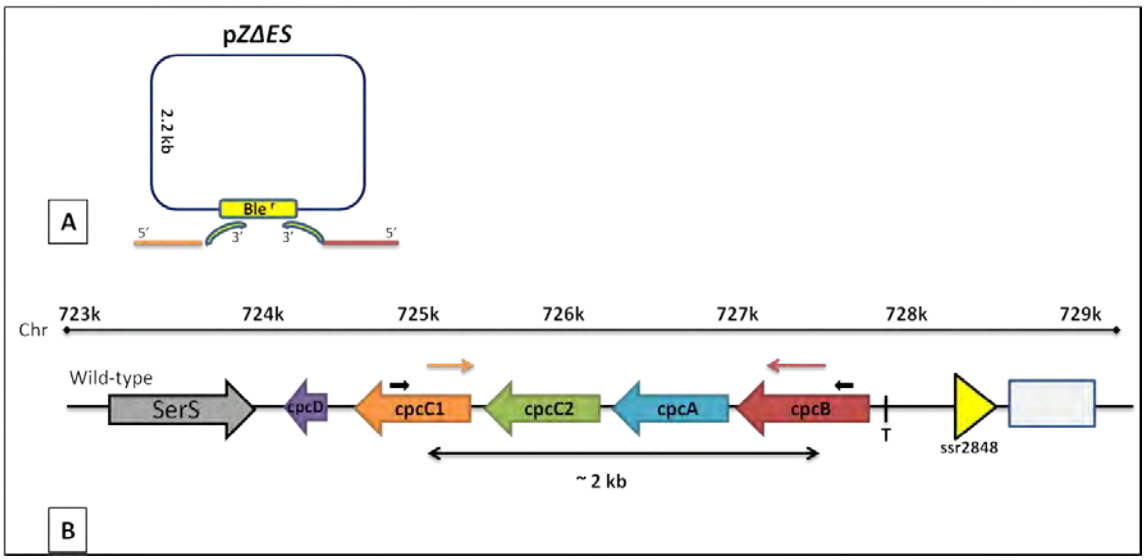


Figure 4.27: Schematic representation of the large chromosomal section proposed for one-step PCR deletion.

(A) pZΔES plasmid was used as template to amplify a deleted operon using *ble* as the selectable marker and using long-tail primers (indicated with blue curves) that represent the part of the primer homologous to the selectable marker while parts homologous to the target genes are highlighted in orange and red arrows. (B) 2 kb of a chromosome segment including the full or partial sequence of four genes (*cpcC1, cpcC2, cpcA* and *cpcB*), to be replaced with the *ble* selectable marker (~0.6 kb) in a single PCR step. Figure modified from (Lea-Smith et al., 2013).

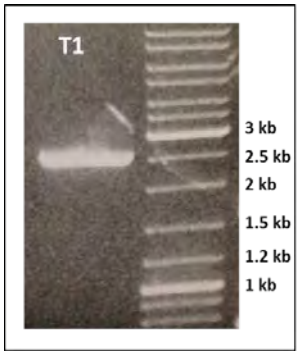


Figure 4.28: PCR result of large chromosomal deletion using one-step PCR.

The one-step PCR method was not successful in deleting the operon consisting of the four genes as shown by the WT band size of 2.5 kb obtained from PCR analysis.

4.4 Conclusion and future work

Work done by several researchers suggests that the minimum length of flanking regions necessary for a successful double homologous recombination in *Synechocystis* is in the region of 200-500 bp. However, results from the first part of this study showed that the minimum length necessary to create targeted gene knockouts using PCR products was as little as 30 bp on either side of the foreign DNA. Nevertheless, there was an obvious difference in transformation efficiency between *Synechocystis* transformed with PCR products with long and short flanks with higher transformants obtained with increased flank lengths (Figure 4.5.A). The information obtained was used to test the possibility of creating targeted gene knockouts in *Synechocystis* using the proposed one-step PCR method. After several unsuccessful attempts to generate targeted gene knockouts of *Synechocystis* using the one-step PCR method using the km^R selectable marker (with a few exceptions), it was found that the km^R cassette tended to insert randomly or in "hot spots" of the *Synechocystis* genome. Sequence results show that nucleotide mutations/deletions and insertions of parts of the km^R cassette in the PCR product obtained using long-tail primers in the one-step PCR may have led to less efficient homologous integration, since the flanks were probably not recognised by the host as homologous to the gene hence identifying the km^R cassette as a naked km^R cassette (no flanks) and inserting it randomly or in hot spots in the genome. This was confirmed when the results were compared to transformants obtained from the transformation of the "naked" km^R cassette (no flanking regions) in chapter 3 that were also integrating in identified "hot spots" in the genome. PCR results presented in Figure 4.10.B indicated random integration of the km^R cassette in the genome. From such results we concluded that non-homologous recombination events were significantly more likely than homologous recombination using PCR products

obtained from the one-step PCR method and that *Synechocystis* 6803 likely contained a recombination system that could insert such a cassette randomly or in “hot spots”.

Studies done by other researchers showed that DNA concentration (Porter, 1986; Kufryk et al., 2002), length of homologous recombination fragments (Williams, 1988; Labarre et al., 1989), and details in the transformation procedure (Kufryk et al., 2002) can affect transformation efficiency in *Synechocystis* (Zang et al, 2007). In this study, we investigated whether similar factors would also play a role in improving the HR:NHR ratio. Results showed that using ssDNA and various DNA and recipient cell concentrations did not improve the ratio. However, we found that using an improved strain of *Synechocystis* (*recJ*- strain) did result, for the first time, in the isolation of large numbers of targeted *psbN*::km^R (Figure 4.17). However, the results were not reproducible and further studies showed that the choice of the selectable marker played a more significant role in obtaining targeted gene knockouts. This was confirmed when three targeted gene deletions of the genes *psbN, psb28.2* and *psbA2* were obtained using the one-step PCR method only after the km^R marker gene was replaced with the *ble* selectable marker (Figure 4.25). Attempts to create gene knockouts using 30 bases of homology linked to the *ble* marker did not work, nor did the attempts to generate large chromosomal deletions (Figure 4.26).

Finally, we can conclude that a one-step PCR-based method for creating targeted gene knockouts in *Synechocystis* has been developed. However, it appears that the following conditions need to be met for the successful application of this method (i) a *recJ*- strain of *Synechocystis* as opposed to the WT *Synechocystis* strain needs to be used as the recipient for transformation (ii) the *ble* marker should be used as the selectable marker of choice (use of the km^R cassette should be avoided) and finally (iii) at least 50 bases of homology to the target gene needs to be used.

Recommended future work would include creating more targeted gene knockouts using the *ble* marker. This will confirm the efficiency and reproducibility of the developed method. Furthermore, checking the use of the developed method using a range of selectable markers (*e.g.* streptomycin, gentamicin and chloramphenicol resistance markers) would allow for the development of a list of “recommended markers” to be used with this method.

Chapter 5: The potential of engineering *Synechocystis* 6803 for photosynthetic terpene production

5.0 Introduction

Terpenes, also referred to as terpenoids and isoprenoids, are typically derived from plant sources and have many industrial applications. Section 1.4.2 in the Introduction chapter describes the industrial applications of these classes of biological compounds, most notably in the perfume industry, pharmaceuticals and as insect repellants (Sikkema et al., 1995, Dunlop et al., 2011, Ro et al., 2006). Monoterpenes, such as geraniol, are commonly used as a fragrance component in consumer products and as an anticancer and antimicrobial agent (Gilpin et al., 2010, Sato et al., 2007, Andoğan et al., 2002) and its commercial production is highly desirable. On the other hand, there has been a recent hyperbole on terpene-derived fuels due to the many advantages they offer. For example, terpenes contain high energy densities and low freezing points and since they generally have methyl branches, double bonds and ring structures, their fluidity is improved at lower temperatures (Lee et al., 2008). Furthermore, they are considered 'drop-in' or 'fungible' fuels as they are compatible with existing engine infrastructure (Lu, 2010). Most of these terpene-derived fuels however have been produced in genetically engineered heterotrophic hosts such as *E. coli* and yeasts (Peralta-Yahya and Keasling, 2010, Peralta-Yahya et al., 2012, Lee et al., 2008, Atsumi et al., 2009). If similar metabolic engineering approaches were successfully applied to photosynthetic micro-organisms, such as eukaryotic microalgae or cyanobacteria, it would result in an efficient biological system for direct light-driven conversion of CO_2 into high value products and advanced biofuels. To date, there are only two reports of isoprenoid pathway engineering in *Synechocystis* with the products being isoprene (Lindberg et al., 2010), and the sesquiterpene, β-caryophyllene (Reinsvold et al., 2011).

In this chapter, research is described that aims at creating transgenic *Synechocystis* strains that have the potential to produce (i) the medium chain (C_{10}) monoterpene geraniol that can serve as both a high value product and a potential biofuel and (ii) the C_{15} sesquiterpene farnesene, a potential diesel replacement fuel. This was to be achieved by means of genetic engineering techniques to introduce novel terpene biosynthesis enzymes that convert metabolic intermediates to these high value products and engine ready fuels. The chapter also describes toxicity tests of selected terpenes on *Synechocystis* and methods used to mitigate the toxic effects of these terpenes on the cyanobacterium.

5.1 Literature review

5.1.1 Terpene production as high value products and biofuels

Historically, the use of prokaryotes as a source of terpenes was not considered feasible (Daum et al., 2009), possibly as terpenes are naturally produced in such small quantities (Martin et al., 2003). However, this situation has changed dramatically over the past decade with the advancement of metabolic engineering. Due to the many applications of terpenes in industry, engineering metabolic pathways for the production of large quantities of terpenes from amenable biological hosts seems more attractive than chemically synthesizing the terpenes or extracting them from the natural source where their amounts are limited (Martin et al., 2003).

In recent years for example, there has been considerable progress in the use of synthetic biology tools to re-engineer metabolic pathways in the heterotrophic model organism *E. coli* to produce a range of novel hydrocarbons (Peralta-Yahya et al., 2012). Such hydrocarbons including long chain alcohols, fatty acid-derived and terpene-derived fuels, have the potential to be used as fungible fuel molecules and currently, are being developed to serve as either supplements or drop-in replacements for existing petroleum fuels (Dunlop, 2011). One example of a potential terpene-derived fuel is the monoterpene alcohol geraniol which is considered a superior gasoline replacement compared to ethanol (Peralta-Yahya and Keasling, 2010) and a high value product. The production of geraniol has been reported in modified yeast (Oswald et al., 2007) and tomato (Davidovich-Rikanati et al., 2007). Farnesene on the other hand makes a good biodiesel replacement fuel (Rude and Schirmer, 2009) and has been produced in both genetically engineered *E. coli* (Wang et al., 2011) and *S. cerevisiae* (Peralta-Yahya et al., 2012, Brennan et al., 2012).

To date, out of the terpene-based biofuels being produced, farnes**a**ne is the closest to commercialization. This is because farnes**a**ne contains a higher cetane number (58) relative to farnes**e**ne, which translates to higher energy density, and is therefore a preferred diesel fuel replacement. Nevertheless, since farnes**a**ne is produced by means of chemical hydrogenation of farnes**e**ne, it is vital to be able to synthesise this fuel precursor (Rude and Schirmer, 2009).

Indeed, promising initial work has been reported in which the model species, *Synechocystis,* or other genetically tractable cyanobacterial species have been

engineered to produce a range of industrial compounds including biofuels (Ducat et al., 2011b). Table 1.1 in the introduction chapter shows several examples of high value chemicals (fatty alcohols) and potential high-energy fuels (hydrocarbons) that have been produced in various cyanobacteria.

As discussed in chapter one sections 1.4.1 and 1.4.2, the terpenoid (or isoprenoid) pathway in *Synechocystis* is a particularly attractive target for next generation biofuel engineering. The isoprenoid pathway is responsible for the biosynthesis of a wide range of compounds using five-carbon isoprene units as building blocks (Vranova et al., 2012). Figure 5.1 gives an overview of the basic pathway and highlights how the introduction of additional enzymes could create novel C_5, C_{10} or C_{15} products that are potential biosynthetic alternatives to gasoline, diesel, and jet fuel. In many cases, only a single enzyme such as isoprene synthase or geraniol synthase is required, although additional enzymes could be employed to further modify the product. For example, adding geraniol dehydrogenase will convert geraniol to geranial, which is another potential biofuel. A desirable objective will be to extend the isoprenoid pathway of *Synechocystis* for synthesis of monoterpenes and sesquiterpenes that can be used as potential advanced biofuels.

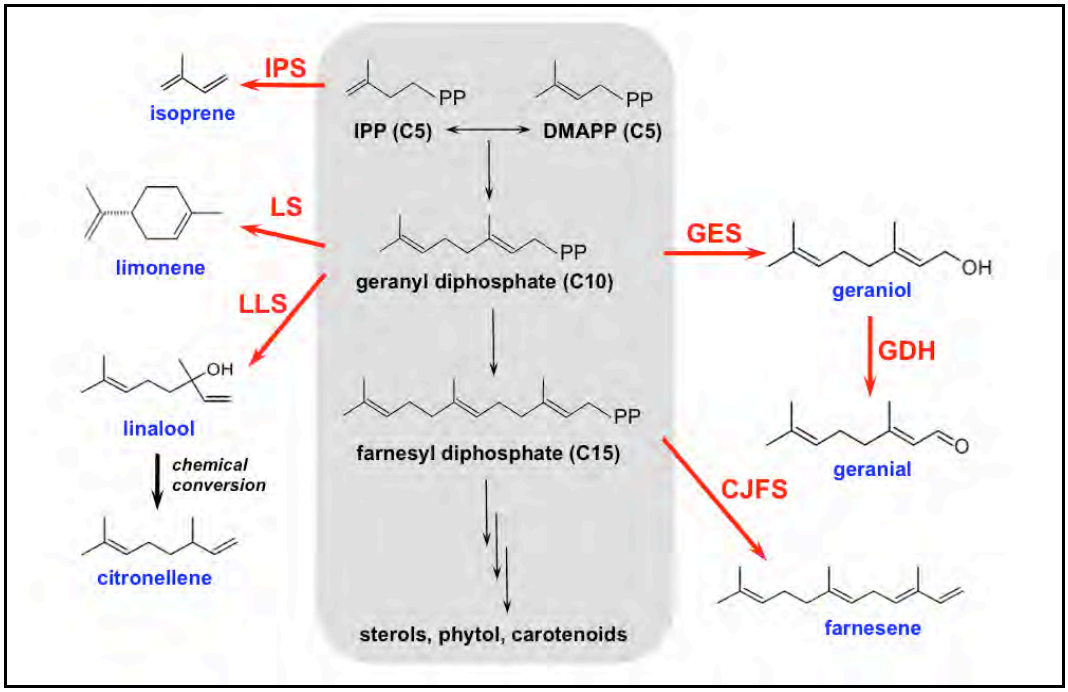


Figure 5.1: The isoprenoid pathway in *Synechosystis*

The isoprenoid pathway in *Synechocystis* is highlighted in grey. The pathway gives examples of possible pathways (red) that could be introduced to produce novel hydrocarbons (blue). **IPP**=isopentenyl diphosphate; **DMAPP=** dimethylallyl diphosphate; **IPS**: isoprene synthase; **LS**: limonene synthase; **LLS**: linalool synthase; **GES**: geraniol synthase; **GDH**: geraniol dehydrogenase; **CJFS**: (E)-β-farnesene synthase.

5.1.2 Toxicity of terpenes on micro-organisms

Despite the fact that advanced biofuels such as terpene-derived fuels may offer many advantages as mentioned earlier, they are often toxic to the microbial host (Andoğan et al., 2002, Azevedo et al., 2012, Dunlop, 2011, Somolinos et al., 2010). Product toxicity introduces an undesirable trade off in engineered strains since the producing strain must balance production against survival. This is a common problem in strain engineering for biotechnology applications (Dunlop et al., 2011). The level of tolerance of microorganisms to advanced biofuels varies amongst different species and strains and also depends on the type of biofuel being produced. Studies on the mechanisms of toxicity of terpenes on microorganisms indicate that the key factor that contributes to the toxic effect correlates to the hydrophobicity (also known as lipophilicity) of the produced terpene (Dunlop, 2011). Hydrophobicity determines the extent by which the terpene accumulates on the cytoplasmic membrane and therefore how well it partitions in the cell membrane. This is often measured by the octanol-water partition coefficient (log $P_{ow)}$. Compounds with log P_{ow} values between 1 and 5 are highly toxic for microorganisms (Dunlop et al., 2011, Isken and de Bont, 1998, Sikkema et al., 1995).

It is essential to understand the mode of toxicity of a synthesised terpene in order to be able to mitigate its effect. The mechanisms of membrane toxicity have been extensively reviewed (Dunlop, 2011, Sikkema et al., 1995) and can be summarised as follows: The accumulation of hydrophobic compounds in the cytoplasmic membrane results in an increased membrane permeability and fluidity which ultimately leads to the release of ATP, ions, RNA, phospholipids and protein. This in turn will lead to diminished energy status of the cell, loss of membrane protein function and affect the overall stability and structure of the cell membrane and its overall role as a barrier. It is therefore crucial to improve tolerance in parallel with the development of genetically engineered strains for the production of advanced biofuels.

The development of methods that improve tolerance of microorganisms to such products has gained a lot of interest. In recent years, a range of approaches has been described (Brennan et al., 2012, Dunlop, 2011, Dunlop et al., 2011, Isken and de Bont, 1998, Sikkema et al., 1995). These include expressing efflux pumps, heat shock proteins, membrane modifying proteins and activations of general stress response genes. The detailed description of these methods is

beyond the scope of this chapter and can be found in Dunlop, 2011. Another method that can help mitigate the toxic effect of advanced biofuels is controlled production by means of a tightly regulated promoter. In this method, the production of biofuels would be switched on only after the initial growth phase has been established. A method that has been reported to be successful in alleviating biofuel toxicity (Brennan et al., 2012).

An alternative to employing molecular engineering of cyanobacteria for improved tolerance is to use a biphasic extraction system whereby an inert solvent phase is added to the medium to create an isolation layer that helps separate the biofuel from the producing host. This has been shown to alleviate terpene toxicity in *Saccharomyces cervisiae* (Brennan et al., 2012). This chapter will focus on attempts made to use both an inducible promoter system to achieve controlled terpene production and a two-phase extraction method for improved biofuel tolerance.

5.1.3 Physical properties and engine performance of advanced biofuels

Identified metabolic pathways offer several possible avenues for the biosynthesis of novel fuel molecules (Fortman, 2008) but before producing such fuels in an industrial setting, the physical and chemical properties of the potential biofuel should be considered. These properties impact on almost everything from suitability as a fuel for extraction, purification, processing, to their behavior in combustion engines. There are various reviews that emphasise the importance of considering factors such as engine type (spark or compression ignition), energy content, combustion quality or ignition delay, cloud point, volatility, lubricity, viscosity, stability, odour, toxicity, water miscibility and cost when developing advanced biofuels from micro-organisms (Antoni et al., 2007, Lee et al., 2008, Peralta-Yahya and Keasling, 2010). The compatibility of the biofuel with the engine type is a very important factor. It is often that engines have to either be adapted to the biofuel or the biofuel has to be designed to mimic, as closely as possible, all important features of traditional fossil fuels (Antoni et al., 2007). Alternatively, biofuels could be blended with traditional fuels.

It is essential, when developing future fuels, to make use of advances in many fields of science and engineering to ensure that such fuels are sustainable in both production and utilization (Hellier et al., 2013). To date, and to the best of my knowledge, there has not been a study combining the microbial production of

advanced biofuels together with experimental studies on an engine. Attached to this thesis is a joint publication that was an outcome of the collaborative work with Dr Paul Hellier from the Department of Mechanical Engineering at University College London where some of the engine parameters listed above was discussed alongside the possible biological production of selected terpenes in *Synechocystis*. The initial idea was to test the diesel engine performance of the C_{10} compound, geraniol, to explore its potential as a possible diesel engine replacement. However, 11 other terpenes that represent a change to the molecular structure of geraniol were also selected for the investigation. This was for the personal interest of Dr Paul Hellier whom conducted the engine tests.

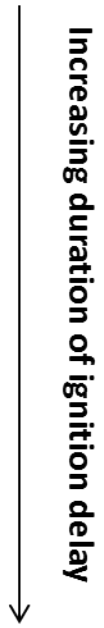
In summary, conclusions as to the suitability of the terpenes as potential single component fuels in diesel engines were, in general based on the ignition delay time of the tested fuel and a ranking was produced to show best to worse diesel replacement fuel as seen in Table 5.1. By virtue, what is a poor diesel fuel would potentially prove better suited to be a gasoline fuel and vice versa (Hellier et al., 2013).

Table 5.1: Ranking of 12 selected terpenes from best to worst performance in diesel engine.

Ranking was done according to the ignition delay and emissions produced.

Test diesel fuel

- Reference fossil diesel
- Geranial (Citral-a)
- Farnesene
- Squalene (produces excessive amounts of soot)
- Citral dimethyl acetal
- 3, 7 - dimethyl - 1 - octanol
- Nerol
- Geranyl acetate
- Citronellol
- Geraniol
- L-Menthol
- Citronellene
- Linalool



5.2 Chapter background and aims

In the work presented here, the aim was to genetically engineer *Synechocystis* to generate transgenic strains capable of geraniol production mainly as a potential high value product and secondarily, as a possible fungible fuel in conventional diesel or gasoline engines. The toxicity of geraniol on *Synechocystis,* and the effect of a two-phase culture system on *Synechocystis* tolerance towards geraniol were investigated. Furthermore, the performance of geraniol along with other selected terpenes as fuels for diesel engines were assessed in collaboration with Dr Paul Hellier from the Department of Mechanical Engineering at UCL. Results from engine tests suggested other terpenes (*e.g.* farnesene) as superior fuels, and these were further explored through toxicity testing and additional transgenic lines were created.

5.3 Results and discussion

5.3.1 Constructing and testing the pLAH.A2 and pLAH.nrsB expression vectors

In order to transform *Synechocystis* it was essential to build and test vectors that would allow for the successful integration and expression of foreign genes in *Synechocystis*. For this, the pLAH.A2 and the pLAH.nrsB plasmid vectors were constructed and tested. In both constructs the *psbA2* gene (gene code = slr1311) was selected as the site of insertion of the foreign genes. The *psbA2* gene is one of three genes (*psbA1*, *psbA2* and *psbA3*) that codes for the D1 protein of photosystem II (Vermaas, 1998) that is believed to be responsible for protection against photodamage or stress (Ikeuchi and Tabata, 2001). The reason behind selecting the *psbA2* locus is that it has been shown that *psbA2* knockouts can grow photoautotrophically given that the other genes are still active (Mohamed et al., 1993) and its deletion does not lead to a reported phenotype (Vermaas, 2007). The following sections details the construction of the two expression vectors and experiments conducted to test their functionality.

5.3.1.1 Constructing the pLAH.A2 expression vector:

The pLAH.A2 vector was constructed to allow expression of foreign genes under the control of the native constitutive (but up-regulated by light) *psbA2* promoter.

The genome sequence of *Synechocystis* was consulted through Cyanobase [8] to design primers to amplify a 2530 bp sequence around *psbA2* from genomic DNA. This amplified DNA was then used as a template to amplify both the upstream and downstream elements of the coding region used to construct the integration platform. Two regions of the *Synechocystis* genomic DNA containing 574 bp and 723 bp of sequence located immediately upstream and downstream, respectively, of the *psbA2* coding region were amplified to give elements with suitable restriction enzyme sites at either end and then assembled into a cloning vector using classical molecular techniques. The primers used are listed in Appendix 3. Finally, a Tn903 kanamycin resistance cassette from pUC4K was digested with *Hinc*II before cloning it in the *Sma*I site of the expression vector resulting in the complete assembly of the 4450 bp pLAH.A2 expression vector as showed in Figure 5.2.

An *Nde*I site was introduced at the translation start to allow for a simple in-frame introduction of any gene into the integration platform under the control of the *psbA2* promoter. Similar *psbA2*-based vectors have been described (Lagarde et al., 2000).

[8] http://www.kazusa.or.jp/cyano/cyano.html

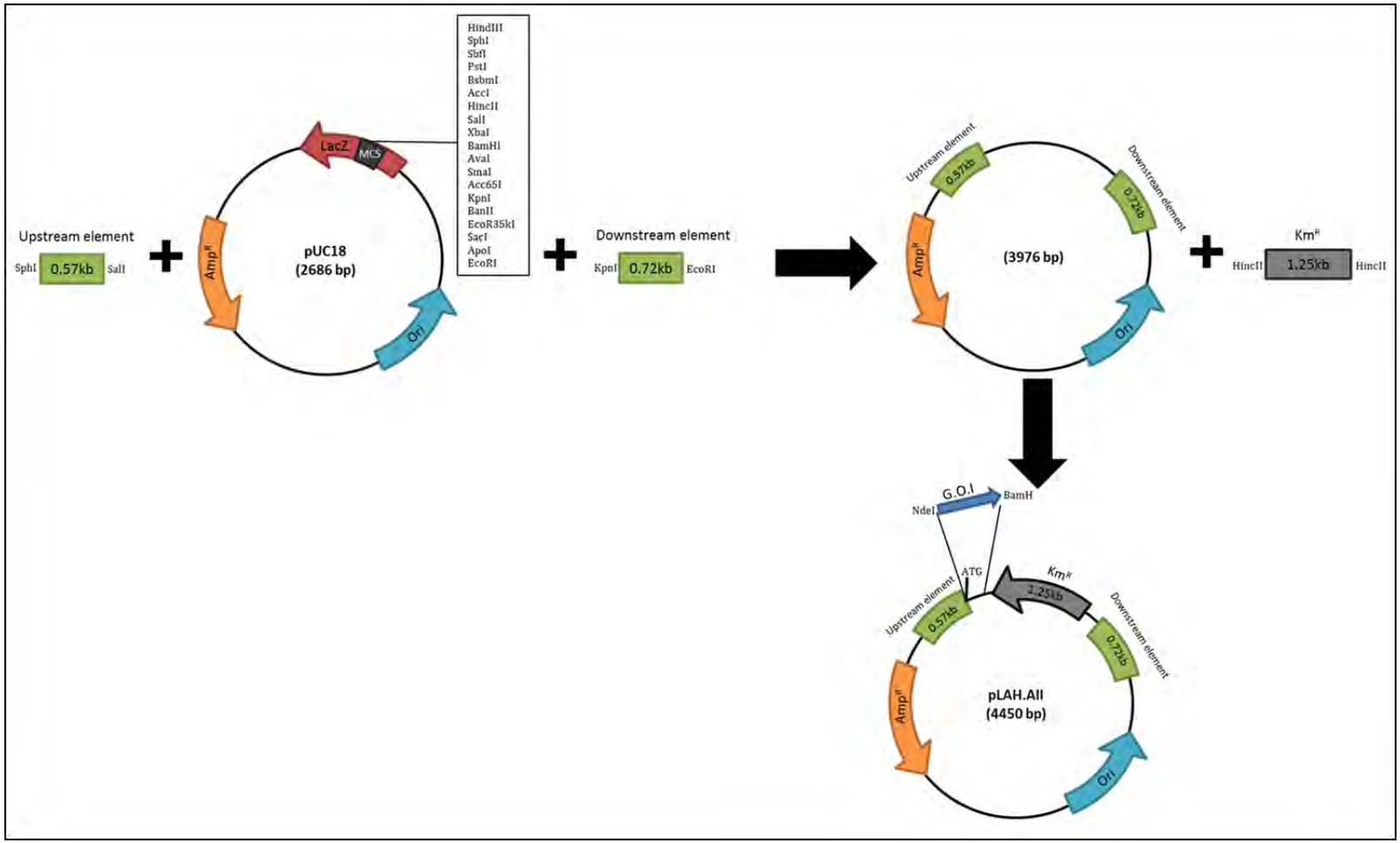


Figure 5.2: pLAH.A2 expression vector including the psbA2 light regulated promoter.

Key: Green rectangles represent the upstream (*including the psbAII* promoter sequence) and downstream *psbAII* elements, **MCS**; multiple cloning site, **Amp^R**; ampicillin resistance caseett, **Ori**; origion of replication, **km^R**: kanamycin resistcane cassett, **G.O.I**; gene of interest.

5.3.1.2 Constructing the pLAH.nrsB expression vector:

The pLAH.nrsB vector was constructed to allow expression of foreign genes under the control of the native tightly controlled, nickel inducible promoter (*nrsB*). The *nrsB* promoter is part of the *nrs*BACD operon that encodes four proteins involved in Ni^{2+} resistance and upstream of this operon are the *nrs*R and *nrs*S genes which make the two-component signal transduction system involved in Ni^{2+} sensing and induction of the operon (López-Maury et al., 2002). In this expression vector the *nrsB* promoter along with its regulatory element were used to drive the expression of foreign genes.

Once more, the genome sequence of *Synechocystis* was consulted through Cyanobase to design primers needed to amplify a 634 bp region of the *Synechocystis* genomic DNA located immediately upstream of the *psbA2* coding region (avoiding the *psbA2* promoter) to give an element with *Hind*III enzyme sites at both ends. The upstream element was then cloned in the *Hind*III site of the pUC18 already containing the same downstream element as that in the pLAH.A2 expression vector. A km^R resistance cassette was isolated from pUC4K plasmid by digesting it with *Hinc*II and subsequently cloned in the *Sma*I site of the expression vector. The 875 bp *nrsB* promoter sequence along with its regulatory element (*nrsRS*) was amplified using primers that added *Sph*I and *Xho*I restriction sites at either end and an *Nde*I site used for the start of transcription on the gene of interest. Primers used are listed in Appendix 3. The *nrsB* sequence was digested with *Sph*I and *Xho*I and subsequently cloned into the *Sph*I and *Sal*I sites of expression vector (*Xho*I and *Sal*I generate the same overhangs). This resulted in the complete assembly of the 5389 bp pLAH.nrsB expression vector as showed in Figure 5.3.

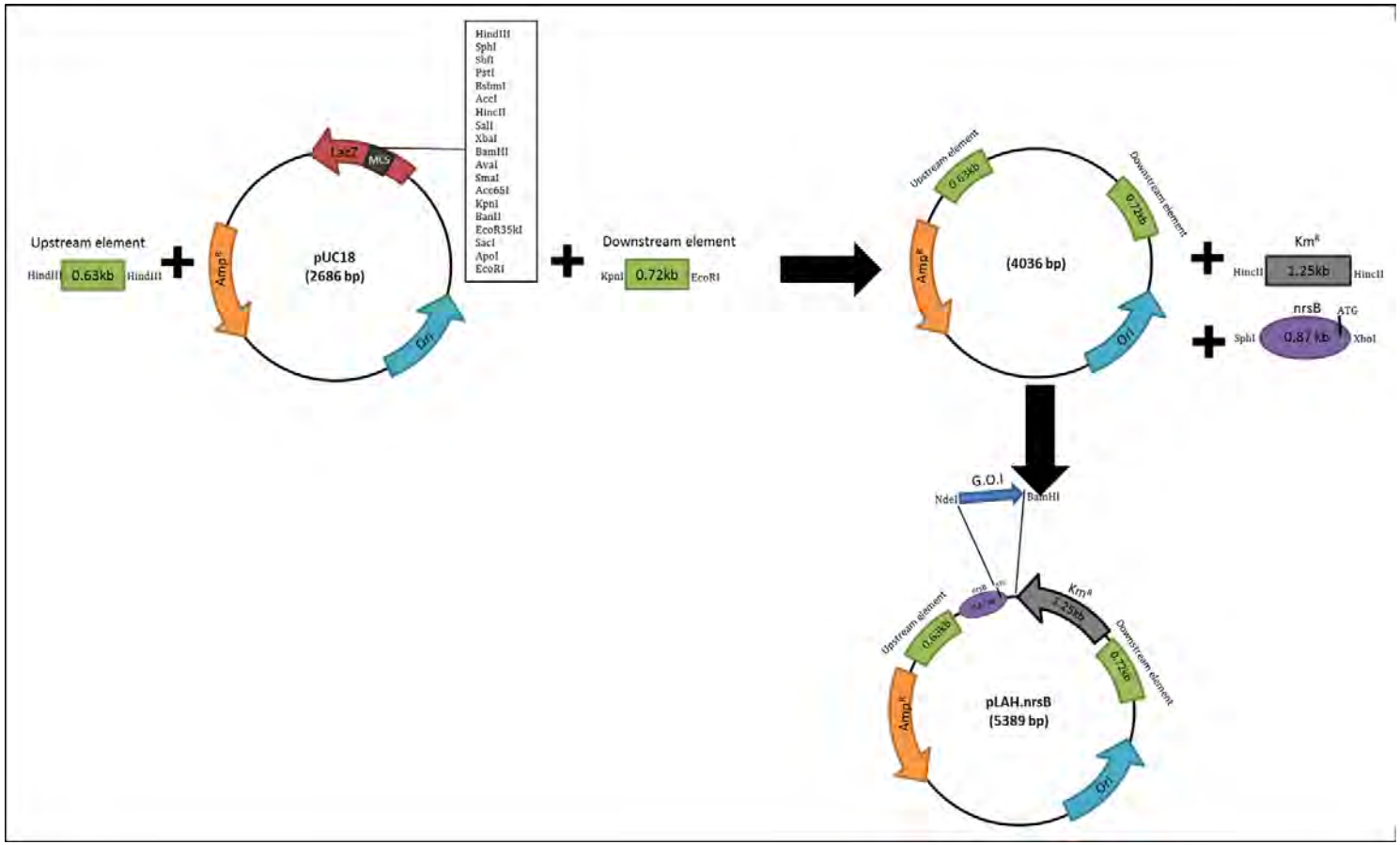


Figure 5.3: pLAH.nrsB expression vector including the nrsB inducible promoter.

Key: Green rectangles represent the upstream (excluding the *psbAII* promoter sequence) and downstream *psbAII* elements, **MCS**; multiple cloning site, **Amp^R**; ampicillin resistance cassette, **Ori**; origin of replication, **km^R**: kanamycin resistance cassette, **G.O.I**; gene of interest, **nrsB**; inducible promoter including its regulatory element.

5.3.1.3 Testing the functionality of pLAH.A2 and pLAH.nrsB expression vectors

The two expression vectors described in previous sections were constructed to allow the expression of selected terpenes in *Synechocystis.* In case of the expression of non-toxic or less toxic terpenes, the pLAH.A2 expression vector allows expression of genes constitutively under the control of the *psbA2* promoter. The pLAH.nrsB expression vector allows for the expression of more toxic genes under tightly controlled inducible conditions. This section describes results from experiments conducted to test the functionality of the two expression vectors.

The *ble* selectable marker conferring resistance to zeocin was used to test the two expression systems. Details of primers used to amplify the 381 bp *ble* sequence from the pRecJ:ble plasmid (Appendix 1) with engineered *Nde*I and *BamH*I sites can be found in Appendix 3. This product was cloned into *Nde*I and *BamH*I sites of the two expression vectors using standard molecular techniques creating the pLAH.A2.ble and pLAH.nrsB.ble plasmids. The latter was sent for sequencing and results confirm the correct cloning of the *ble* marker under the control of the *nrsB* promoter (Appendix 9).

Synechocystis was then transformed with the two plasmids as described in section 2.4.2. Putative transformants appeared within seven days post selection on medium supplemented with zeocin at 25 µg/ml and in addition to 6.4 µM of $NiSO_4.6H_2O$ for transformants expressing the gene under the *nrsB* promoter. There were then taken through three rounds of selection on zeocin-containing medium to ensure they reached homoplasmy before genomic DNA was isolated (section 2.3.3.1) for PCR analysis (section 2.3.5). The correct integration of the *ble* marker into the genome of *Synechocystis* at the *psbA2* site under the control of the two promoters was confirmed by PCR, with results for the *nrsB* plasmid shown in Figure 5.4, using primers located within the upstream element of the *psbA2* gene and the other within the km^R cassette. Details of primers can be found in Appendix 3.

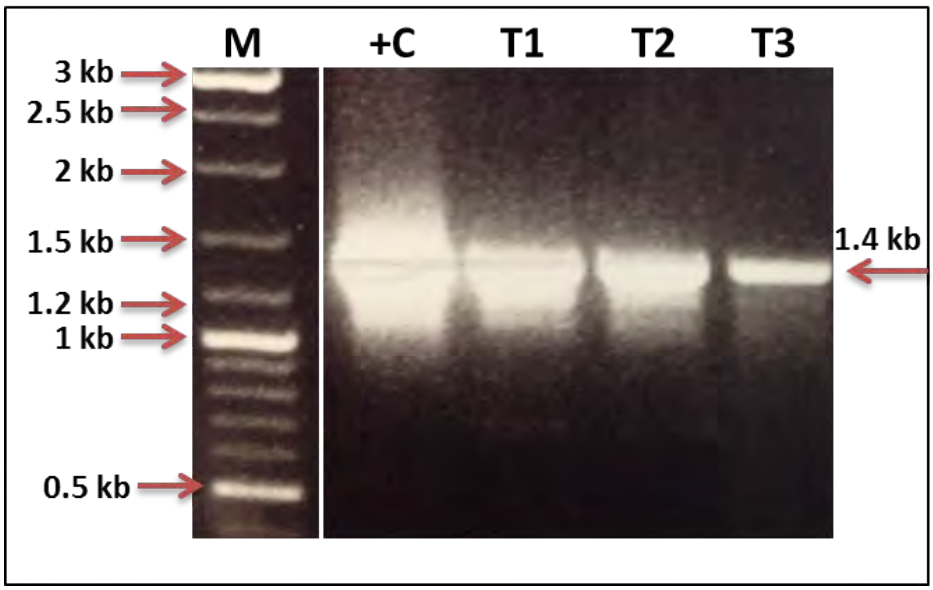


Figure 5.4: PCR results confirming the integration of the ble marker gene in the psbA2 locus.

Three isolated transformants show the presence of *ble* marker gene under the *nrsB* promoter (expected band size of 1.4 kb). Positive control (+C): pLAH.nrsB.ble plasmid, M: 1 kb molecular marker, T1-T3; transformants.

The positive transformants were then tested for active and controlled *ble* expression in *Synechocystis* using western blot analysis and spot tests. In both methods, the following were tested: WT *Synechocystis*, a single transformant containing the *ble* selectable marker under the control of the *psbA2* promoter and three independent transformants with *ble* under the *nrsB* promoter (Figure 5.4). Cultures were grown to an OD_{750} of ~0.4. Each culture was then split in two flasks, where one culture was induced with $NiCl_2.6H_2O$ added to a final concentration of 6.4 µM and the other kept un-induced. The cultures were incubated for three hours under standard growth conditions (section 2.2.4) and later used for western blot analysis and spot tests.

Western blot analysis: the western blot was set up according to methods described in section 2.6. The controlled expression of the ble protein, in the three strains expressing *ble* under the inducible promoter, was confirmed by the presence of the protein (~13.6 kDa) in cultures that have been pre-induced with nickel ions added to a final concentration of 6.4 µM (Figure 5.5 lanes 3, 5 & 7). Un-induced cultures showed no ble protein (lanes 4, 6 & 8) confirming the tightly controlled expression of *ble* under the *nrsB* promoter. The constitutive expression of ble protein under the *psbA2* promoter was confirmed by the appearance of a band at ~13.6 kDa in lanes one and two where the expression was independent

of nickel induction. The non-specific bands served as loading controls and showed equal loading of samples. It is worth noting that the level of the ble protein is significantly higher when expressed constitutively under the control of the *psbA2* promoter when compared to its induced expression under the *nrsB* promoter.

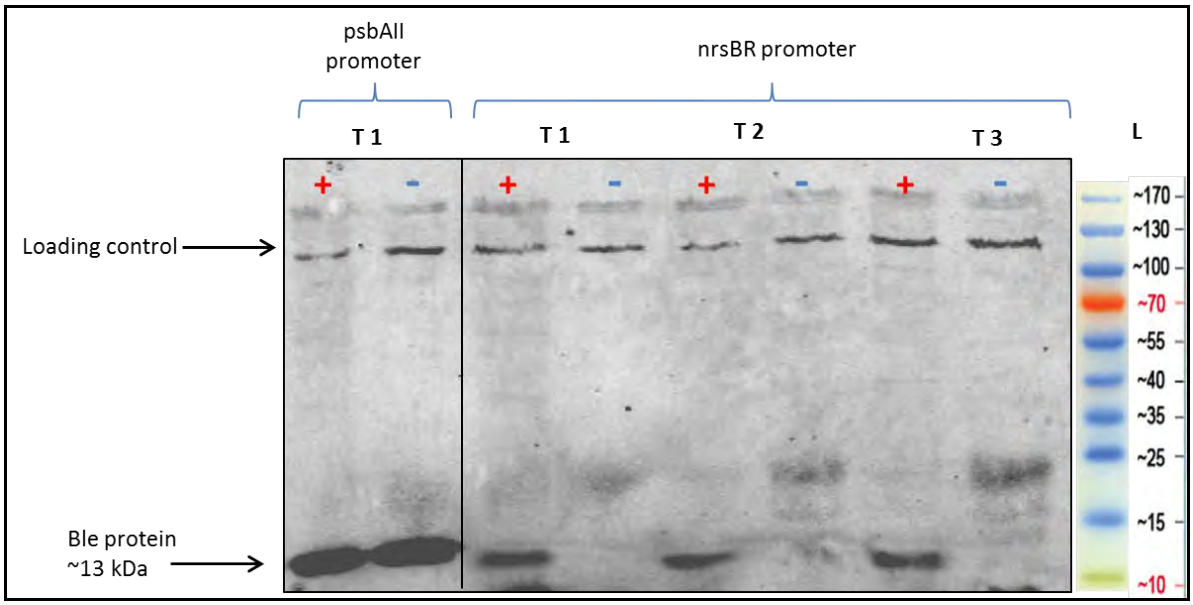


Figure 5.5: Testing the functunality of pLAH.A2 and pLAH.nrsB expression vectrors.

Western blot confirming expression of the ble protein (~13.6 kDa) under constitutive (lane 1 and 2 from the left) and inducible promoter conditions (lane 3, 5 and 7 from the left). (+) indicate cultures induced with $NiSO_4.6H_2O$ while (-) represent un-induced cultures, T1 psbAII and T1-T3 nrsB represent ble expressed under the psbAII promoter and three independent transformants of ble expressed under the nrsB promoter respectively.

The spot test: Growth tests on solid medium ("spot tests") were prepared according to section 2.2.6. The results presented in Figure 5.6 shows the WT and each of the transformants spotted on four plates (done in duplicates) where plate one contained only BG11 medium, plate two contained BG11 supplemented with $NiCl_2.6H_2O$ to a final concentration of 6.4 µM. Plate three contained BG11 supplied with zeocin added to a final concentration of 25 µg/µl and $NiCl_2.6H_2O$ to a final concentration of 6.4 µM. Finally, plate four was supplied with only zeocin to a final concentration of 25 µg/ml.

Plate one, which served as a positive control, showed the healthy growth of the ten cultures of *Synechocystis*. Plate two was used to demonstrate that $NiCl_2.6H_2O$ at concentrations used for induction (6.4 µM) had no detrimental

effect on the cultures and this was confirmed by the healthy growth of the ten cultures. The expression of ble protein under the control of the *psbA2* constitutive promoter was confirmed by the survival of the transformants strain of *Synechocystis* on both plates three and four containing zeocin irrespective of induction. The expression of ble protein under the *nrsB* inducible promoter was confirmed by the growth of the three pre-induced cultures on plate three and the failure of the three un-induced cultures to grow on the same plate demonstrating the expression of ble only under induced culture conditions. However, it is worth noting that when the previously un-induced cultures were plated on BG11 plates containing the inducer ($NiCl_2.6H_2O$) a few colonies appeared on the plates showing the sensitivity of the *nrsB* promoter towards the inducer. The controlled expression of ble under the *nrsB* promoter was demonstrated on plate four where pre-induced cultures that have previously shown to express ble (plate three) failed to grow on plate four containing zeocin in the absence of the inducer. The spot test results confirmed that the inducible system is tightly controlled.

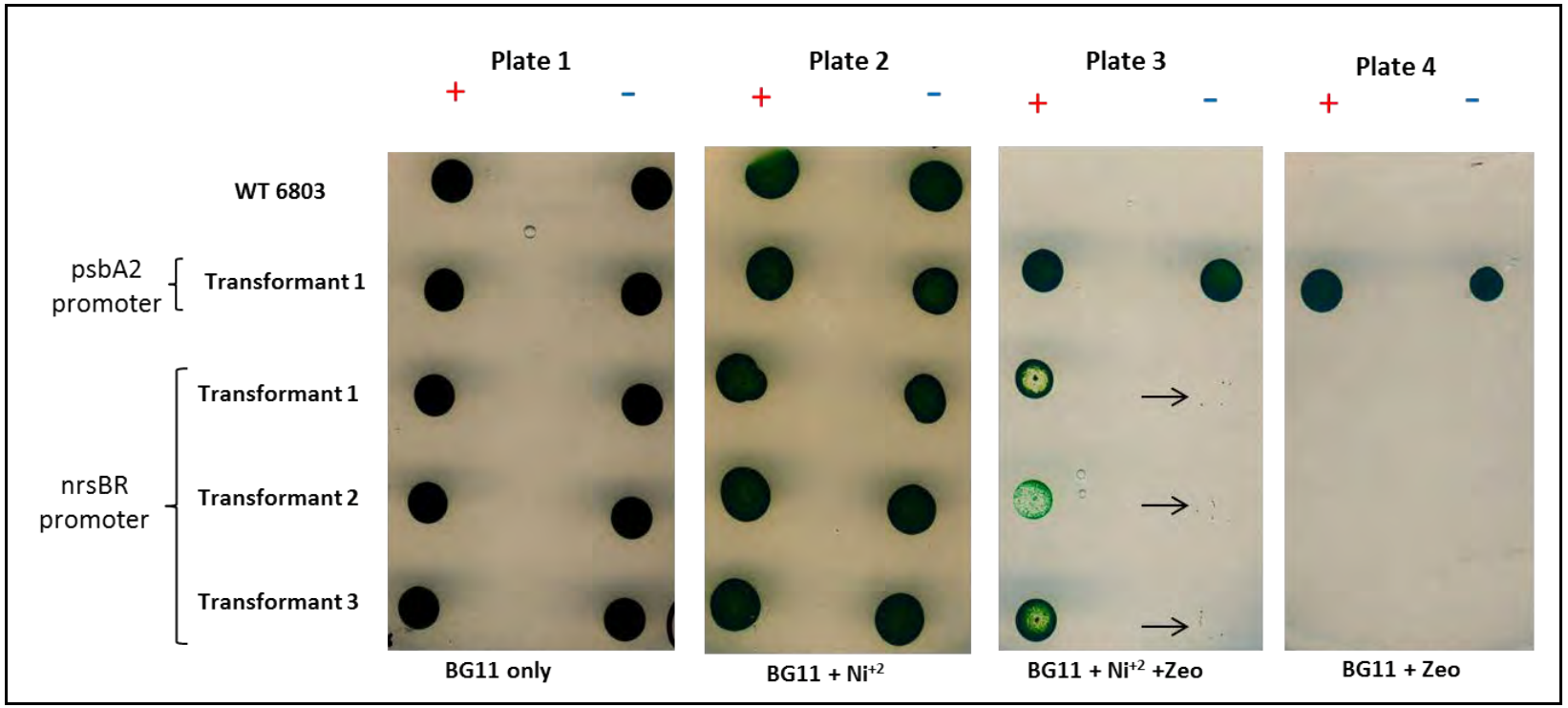


Figure 5.6: Spot tests showing the expression of active ble protein.

Spot tests showing the constitutive and inducible expression of an active ble protein under the *psbA2* and *nrsB* promoters in *Synechocystis*. (+) indicate cultures pre-induced with $NiSO_4.6H_2O$ while (-) represent un-induced cultures.

5.3.2 Potential for geraniol production in *Synechocystis*

Geraniol could potentially be produced in *Synechocystis* by introducing a well characterised Geraniol Synthase (GES) enzyme from sweet basil (*Ocimum basilicum*) into the isoprenoid pathway (Figure 5.1). Geraniol, being a volatile alcohol would diffuse out of the cell upon production thus reducing the economic cost and energy intensity of downstream processing. Furthermore, and as mentioned earlier, the production of geraniol has been achieved in yeast (Oswald et al. (2007), and tomato (Davidovich-Rikanati et al. (2007). All these factors make geraniol an attractive candidate for expression in *Synechocystis*. In this section, the toxicity of geraniol on *Synechocystis* was examined, methods to mitigate its toxic effect were investigated, and attempts to create a transgenic line of *Synechocystis* expressing the *GES* gene are detailed.

5.3.2.1 Geraniol toxicity studies

Toxicity tests were conducted to examine the ability of *Synechocystis* to grow in the presence of increasing amounts of geraniol as described in chapter two (section 2.7.1). The toxic effect of geraniol can be observed visually as seen in Figure 5.7. It is clear that *Synechocystis* cells failed to grow in the presence of even the lowest concentration of geraniol (0.02% v/v) proving that geraniol was extremely toxic when co-inoculated with *Synechocystis*. This result suggested that the constitutive production of geraniol in *Synechocystis* would be difficult to achieve.

On the other hand, an experiment was set up to investigate the levels of tolerance of already established cultures of *Synechocystis* to exogenously added geraniol (section 2.7.2). In this experiment, the tolerance was investigated in liquid cultures by monitoring whether the post addition of geraniol inhibited further growth and led to bleaching of the culture.

As seen in Figure 5.8, adding geraniol to established cultures also displayed a toxic effect indicated by the gradual loss of pigmentation of the cultures over a time period of 72 hours. Cultures inoculated with the lowest concentration of geraniol 0.02% (v/v) survived only 24 hours post addition then gradually lost pigmentation and by 72 hours, were completely bleached (Figure 5.8 A, B and C respectively). Cultures inoculated with 0.1%, 0.2% and 0.4% (v/v) geraniol turned yellow after just 24 hours post inoculation (Figure 5.8 B).

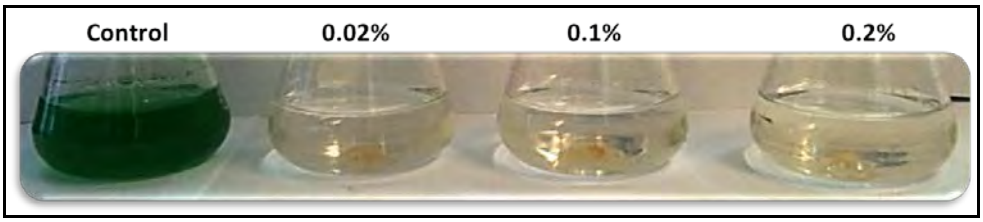


Figure 5.7: Co-inoculation of *Synechocystis* with increasing amounts of geraniol.

Synechocystis failed to grow in the presence of geraniol even at the lowest concentration of 0.02% v/v showing the highly toxic effect of geraniol. Control flask (0% geraniol) showed healthy growth of *Synechocystis* in the culture medium.

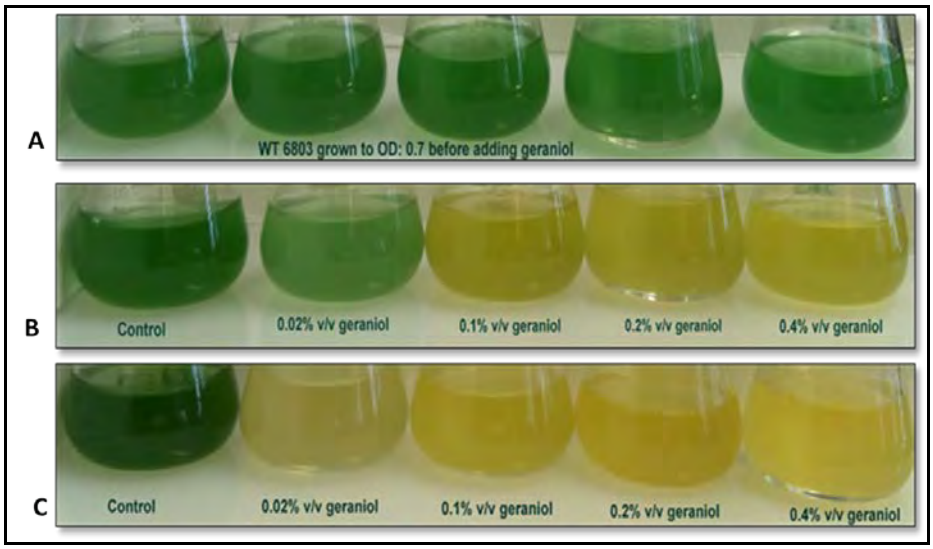


Figure 5.8: Tolerance of established cultures of *Synechocystis* to geraniol.

The tolerance of established cultures of *Synechocystis* to increasing amounts of geraniol 24h and 72h post addition. **(A)** Cultures of *Synechocystis* grown to mid-log phase in the absence of geraniol. **(B)** 24 hours post geraniol addition into established culture of *Synechocystis*. **(C)** 72 hours post geraniol addition into established cultures. Control flask; culture contains 0% geraniol.

To confirm that the loss of pigmentation of *Synechocystis* cultures reflected death of the cells, 200 µl of medium from the cultures shown in Figure 5.8.B were plated on solid medium and tested for cell growth. Only a few colonies (~10) were recovered from cultures inoculated with 0.02% (v/v) geraniol 24 hours post addition and no colonies were recovered from flasks inoculated with 0.1%, 0.2% and 0.4% confirming cell death. In contrast, a lawn of cells was recovered from the positive control (no geraniol) (Figure 5.9). This suggests that even the induced biosynthesis of geraniol in genetically engineered *Synechocystis* would rapidly prove toxic, and thus expression even under these conditions would be challenging. Additional methods would therefore need to be employed to overcome the toxicity issue.

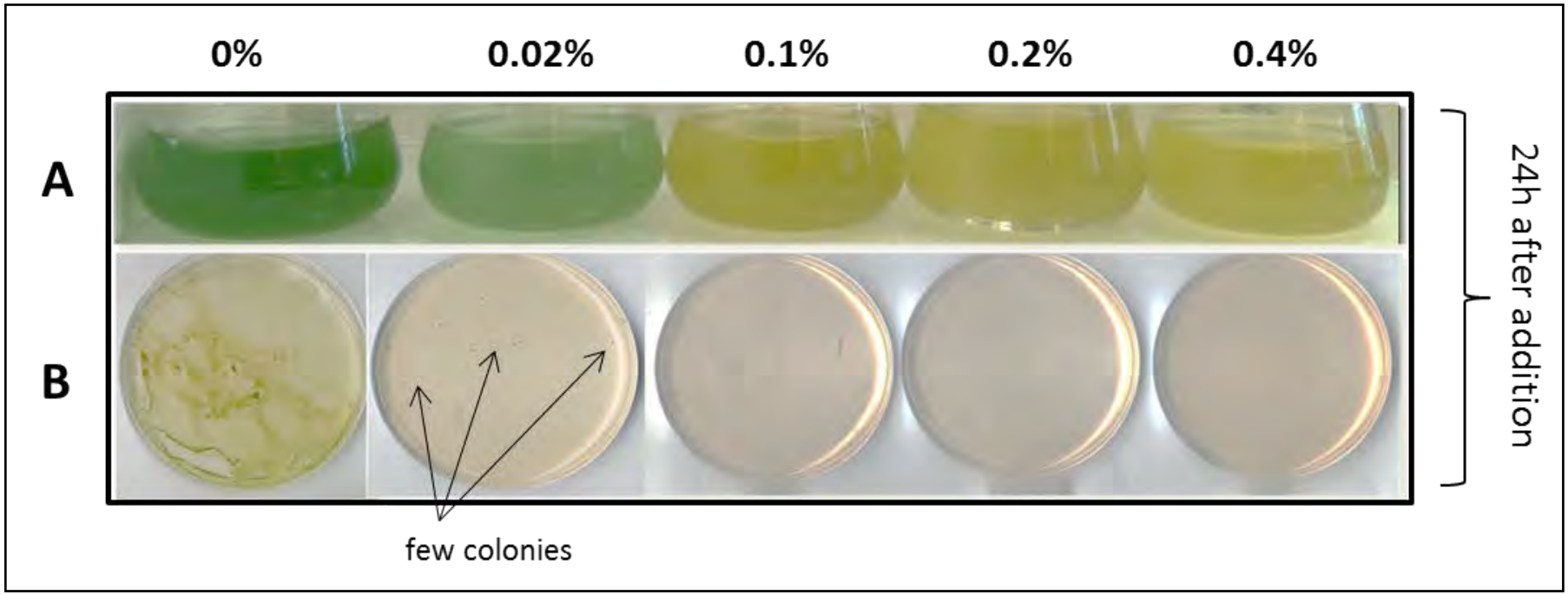


Figure 5.9: Cofirming cell death in treated cultures of *Synechosystis*.

Confirming that death of *Synechocystis* can be represented by the loss of its pigmentation (A) cultures of *Synechocystis* inoculated with increasing amounts of geraniol (%v/v) (B) 200µl of each culture plated on BG11 plates to check cell viability. Full growth appears on the control plate, few colonies from the cultures with 0.02% geraniol while no cells recovered from cultures with higher concentrations of geraniol. Control flask; culture contains 0% geraniol.

5.3.2.2 Effect of a two-phase culture system on tolerance levels

In order to be able to produce geraniol in *Synechocystis,* it was essential to find ways to mitigate the toxic effects of this terpene on the cells. The effectiveness of using a two-phase extraction system to reduce monoterpenes toxicity in *S. cerevisiae* has already been demonstrated (Brennan et al., 2012). The possibility of using a two-phase culture system to potentially "milk" the toxic product from the cultures of *Synechocystis* was therefore investigated. Figure 5.10 shows a simple experiment employing a two-phase culture system, using the hydrophobic solvent n-dodecane, for extraction of geraniol from the culture. The use of n-dodecane as an extracting solvent has already been established (Jang et al., 2011). In this experiment four flasks were inoculated with 25 ml of *Synechocystis* cells grown to an OD_{750} of ~ 0.7. Flask one was used as controls to show normal growth in absence of geraniol and n-dodecane while flask two was used to examine whether n-dodecane itself would have a detrimental effect on the culture. Flasks three and four were used to compare growth of cultures inoculated with 0.02% (v/v) geraniol in the presence or absence of ~17% n-dodecane, respectively. The growth of the cultures was observed over ten days and results are shown in Figure 5.10. Cultures in flask three, containing n-dodecane, showed a marked improvement in the survival of *Synechocystis* even after ten days as compared to cultures in flask four that lacks n-dodecane, (highlighted in yellow rectangle). N-dodecane itself had no detrimental effect on the growth of the *Synechocystis* cultures even after ten days as seen in flask two (similar appearance to cultures in the control flask one). From these results, it can be said, that adding this non-toxic organic solvent to the culture medium helps improve the survival of cells by forming a two phase system in which the toxic product accumulates in the organic phase (indicated by a red arrow in Figure 5.10). This result shows that the controlled biosynthesis of at least small amounts geraniol in *Synechocystis* could potentially be improved significantly in the presence of a media additive such as n-dodecane.

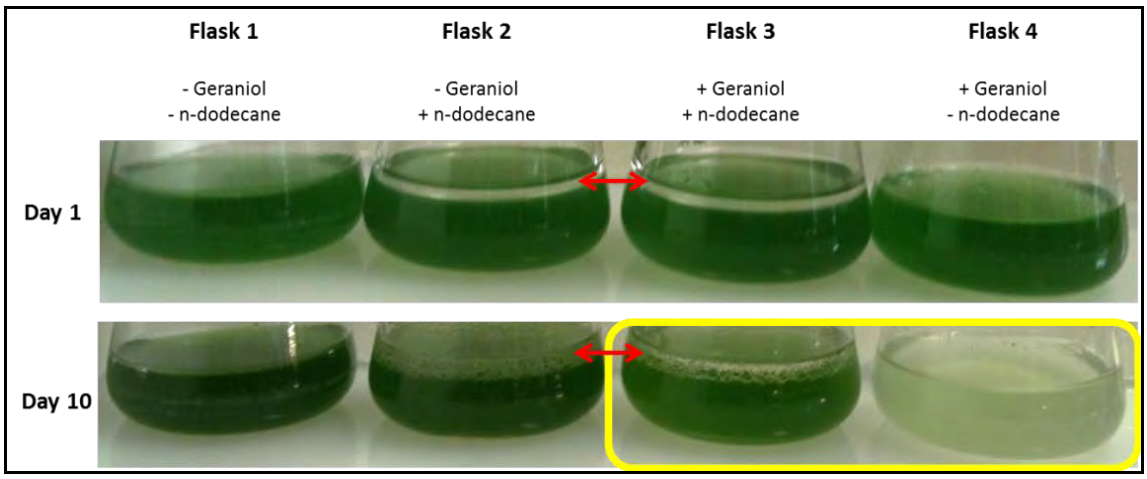


Figure 5.10: Two-phase culture system used to improve tolerance towards geraniol.

Cultures in flask three show a marked improvement in survival of *Synechocystis* in the presence of n-dodecane as compared to cultures grown in the absence of the media additive (flask four).

5.3.2.3 The expression of geraniol synthase in *Synechocystis*

This section describes the isolation and cloning of the cDNA of sweet basil GES into the pLAH.A2 and pLAH.nrsB expression vectors. It also highlights the challenges that were encountered in order to obtain both *E. coli* and *Synechocystis* transformants expressing the *GES* gene under the *psbA2 and nrsB* promoters.

5.3.2.3.i Isolation and cloning of *GES* in the pLAH.A2 and pLAH.nrsB expression vectors.

The pCRT7/CT-TOPO plasmid was used as the source of the sweet basil GES cDNA in these experiments. The cDNA [(encoding a ~ 65 kDa protein (Iijima et al., 2004)] was amplified by PCR without the predicted plastid targeting sequence. The primers used for amplification provided *Nde*I and *Bam*HI restriction sites necessary for cloning of the GES coding sequence (CDS) into the corresponding sites of both pLAH.A2 and pLAH.nrsB expression vectors. Furthermore, an HA-tag sequence was attached to the GES CDS at its 3' end to enable detection using commercial α-HA antibodies. Primers used and sequences of the GES CDS and the pCRT7/CT-TOPO plasmid can be found in Appendices 3, 4 and 5, respectively.

The PCR product of ~1.6 kb was excised from an agarose gel, purified (section 2.3.10) and cloned into the pJET vector. It was then excised using *Nde*I and *Bam*HI enzymes (Figure 5.11) and subsequently ligated into the corresponding

sites of both the pLAH.A2 and pLAH.nrsB expression vectors (section 2.3.11.3). This is depicted in Figure 5.12.

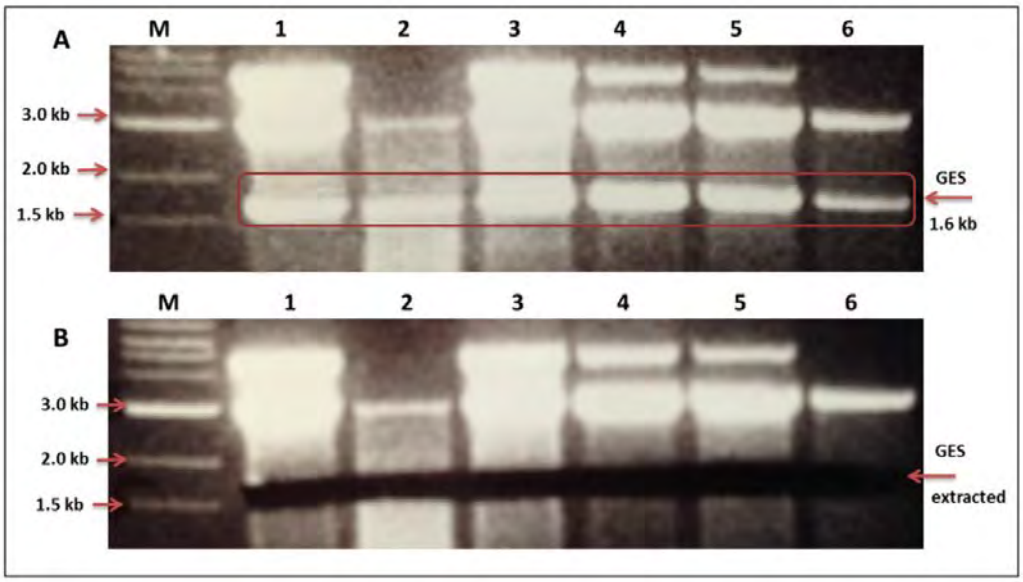


Figure 5.11: *GES* gene isolated from pJET as a *Nde*I and *Bam*HI fragment.

(A) GES CDS (1.6 kb) isolated by double digesting pJET with *Nde*I and *Bam*HI. Lanes 1-6 represent six representative samples of pJET containing the gene. (B) *GES* gene gel extracted and purified for cloning into the pLAH.A2 and pLAH.nrsB expression vectors.

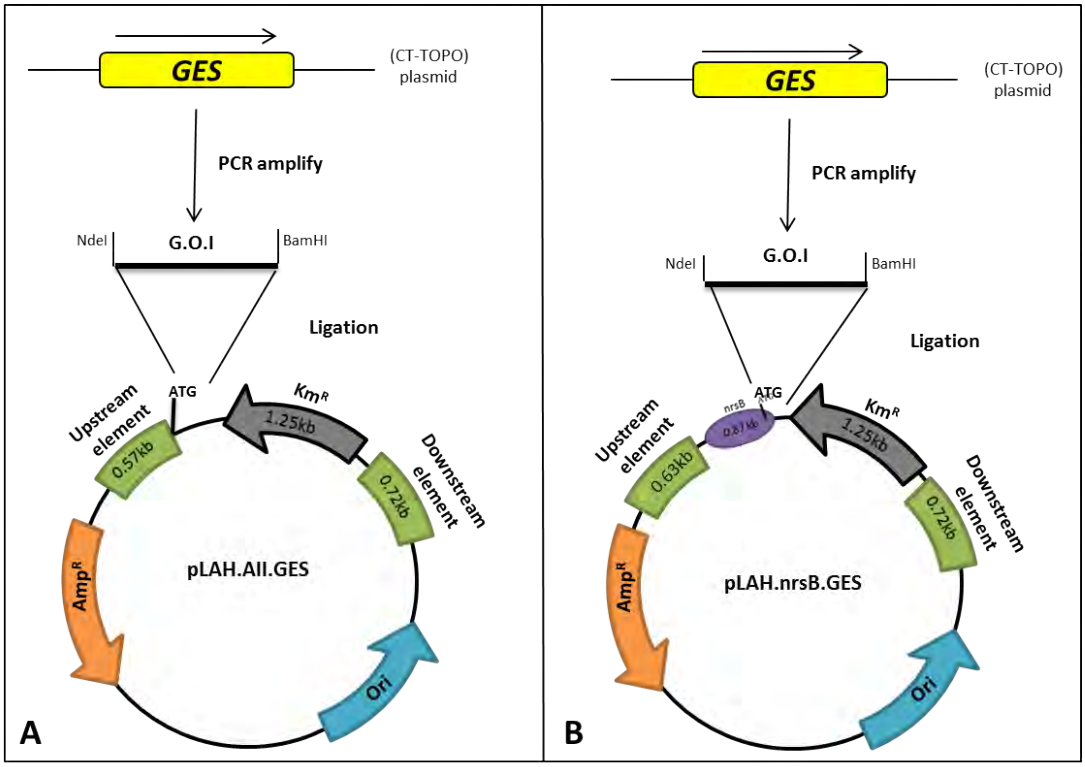


Figure 5.12: A schematic representation showing cloning of *GES* gene under the *psbA2* and *nrsB* promoters.

(A) Cloning steps for the construction of pLAH.A2.GES plasmid. (B) Cloning steps for the construction of pLAH.A2.GES plasmid.

5.3.2.3.ii Transformation of *E. coli*

Transformation of the pLAH.A2.GES and pLAH.nrsB.GES plasmids obtained in section 4.3.2.3.i into competent *E. coli* cells was done according to methods described in section 2.4.1.

The transformation of *E. coli* using pLAH.A2.GES resulted in tens of colonies when selected on LB plates supplemented with 50 µg/ml of kanamycin. However, test digests (using *Nde*I) were performed on plasmids from a total of 20 colonies and revealed the absence of the *GES* gene in the isolated transformants since they showed the same band size as the negative control (Figure 5.13). On the other hand, *E. coli* transformation with pLAH.nrsB.GES resulted in tens of colonies and test digests on six plasmids confirmed the successful isolation of positive transformants indicated by a band shift of 1.6 kb (representing the *GES* gene) from the negative control (result not shown). A single representative plasmid was sent for sequencing and results confirmed the correct insertion of the HA-tagged GES CDS under the *nrsB* promoter (Appendix 13). This plasmid was later used for transformation into WT *Synechocystis* (section 4.3.2.3.iii).

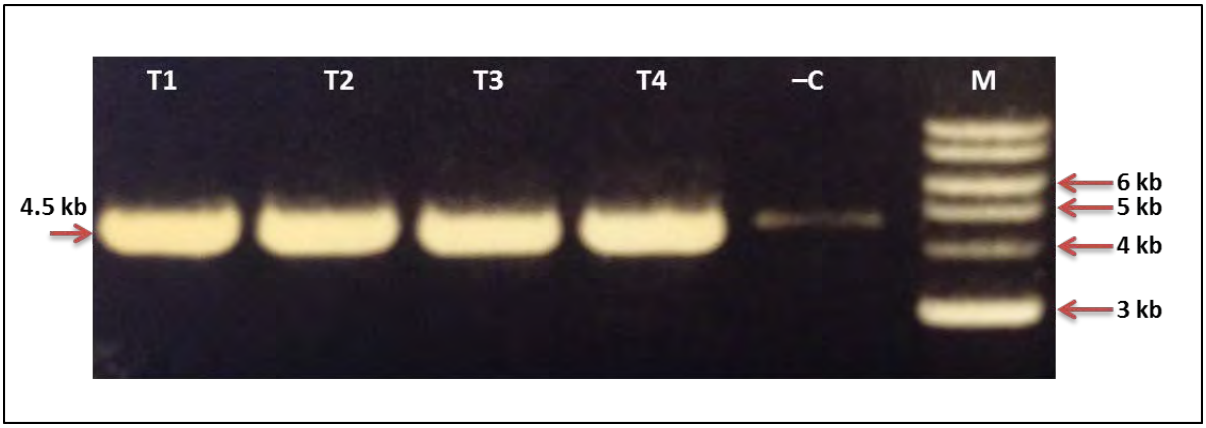


Figure 5.13: Test digests on the pLAH.A2.GES plasmid.

Test digests showed the absence of the *GES* gene in any of the pLAH.A2 clones. T1-T4; transformant lines, negative control: pLAH.A2 expression vector (4.5 kb), M; 1 kb molecular marker.

The failure to recover any pLAH.A2.GES clones was intriguing. According to Eaton- Rye. (2004), *Synechosytsis* PCC 6803 promoters may be recognised by *E. coli* and there are reports of the *Synechocystis psbA2* promoter being functional in *E. coli*. One example is the work done by Utah State University iGEM team in 2010. The team demonstrated a reasonably strong expression of the green fluorescent protein (*GFP*) gene using the *Synechocystis psbA2*

promoter in an *E. coli* host [9]. Moreover, as mentioned in the introduction chapter, there are many studies that describe the antimicrobial activity of geraniol to different microorganisms including *E. coli*. Therefore, it is not unreasonable to conclude that the absence of *E. coli* transformants containing the *GES* gene upon the transformation of *E. coli* with the pLAH.A2.GES was due to the expression of the toxic *GES* gene under the strong *psbA2* promoter that led to cell death due to product toxicity. Furthermore, the fact that positive *E. coli* transformants were obtained upon transforming *E. coli* with the *GES* gene under the *nrsB* promoter further supports this hypothesis. The *nrsB* promoter has not been reported to be functional in *E. coli* and thus there is probably no issue of product toxicity with this promoter. A control experiment was carried out to test whether the expression of GES in *E. coli* under the *Synechocystis psbA2* promoter was the reason behind the absence of *E. coli* positive transformants.

The control experiment: A partially deleted version of the *GES* gene under the control of the *psbA2* promoter was used for transformation into *E. coli* along with the full-length version under the control of the same promoter. Obtaining positive *E. coli* transformants using the deleted version of the *GES* gene and the absence of positive *E. coli* transformants from transformation with the full sequence of the *GES* gene will confirm our hypothesis.

In order to create a non-functional *GES* gene, a “cut and shut” method was used where an enzyme was selected (*Bgl*II) that cuts the GES CDS at two locations without cutting the pCRT7/CT-TOPO plasmid backbone, thereby releasing a 708 bp fragment. The plasmid was re-ligated and PCR was performed on the re-ligated mix using the GES primers as above to give a 0.9 kb product. The pCRT7/CT-TOPO plasmid carrying the full GES CDS (PCR product of ~1.6 kb) was used as a positive control. The PCR products obtained were purified, cloned into pJET before being digested with *Nde*I and *Bam*HI restriction enzymes and ligated into the corresponding sites of the pLAH.A2 expression vector to create the pLAH.A2.GES(F) and pLAH.A2.GES(D) plasmids as described in section 2.3.11.3. These ligations were used to transform competent *E. coli* cells and colonies appeared on LB plates supplemented with kanamycin at 50 µg/ml for both transformations. Plasmids prepared from randomly picked colonies were

[9] http://2010.igem.org/Team:Utah_State

digested using *Nde*I and *Bam*HI restriction enzymes to check for the presence of the deleted and full-length *GES* sequences. As predicted, only the transformation of pLAH.A2 carrying the deleted *GES* gene yielded the correct recombinant plasmid as confirmed by a restriction digest in which a 0.9 kb fragment representing the deleted *GES* gene was observed. On the other hand, none of the ten screened plasmids from the transformation with using the full-length *GES* gave the expected band size of 1.6 kb, with most remaining undigested suggesting the absence of the *GES* DNA. Figure 5.14 shows the results for a representative of each transformant. For confirmation, one of the putative pLAH.A2.GES.F plasmids was sent for sequencing. BLAST results revealed no significant match to GES and that a random piece of DNA was inserted downstream of the *psbA2* promoter (Appendix 10).

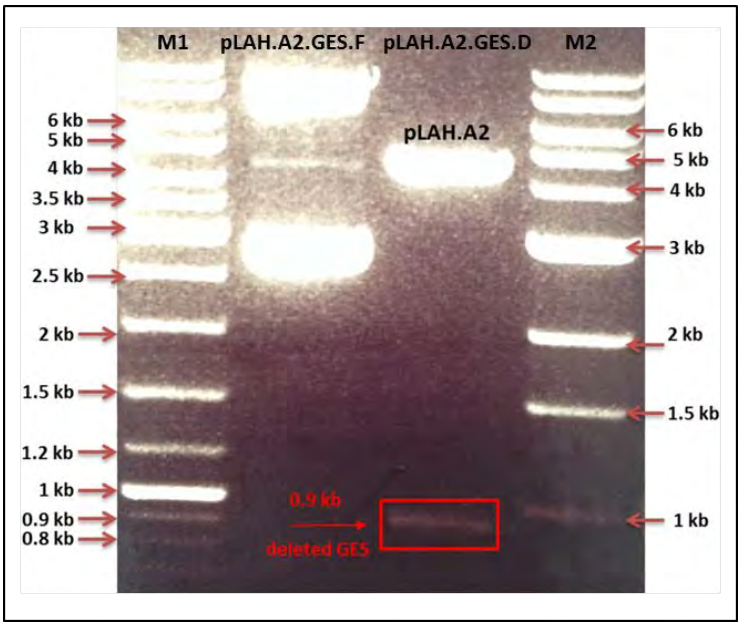


Figure 5.14: Test digests on putative pLAH.A2.GES.F and pLAH.A2.GES.D plasmids.

Test digests confirm the isolation of pLAH.A2.GES.D plasmid by dropping out a 0.9 kb fragment representing the deleted *GES*. The results indicate the absence of the full-length GES in pLAH.A2.GES.F. M1; Fermentase 1 kb molecular marker, M; NEB 1 kb molecular marker.

5.3.2.3.iii Transformation of *Synechocystis*

The pLAH.nrsB.GES and the pLAH.A2.GES.D plasmids obtained in the previous section were used for transformation of WT *Synechocystis* as per methods described in section 2.4.2. For both transformations, colonies appeared within one week of selection on medium supplemented with kanamycin added to a final concentration of 200 µg/ml. Genomic DNA was isolated from several colonies

from each transformation experiment and analysed by PCR using primers designed to check for correct integration of the *GES* gene in the *psbA2* locus.

Unfortunately, PCR results confirm that none of the five representative transformants containing the *GES* gene under the control of the *nrsB* promoter were positive for GES (Figure 5.15.A). To confirm the results obtained in Figure 5.15.A, an additional PCR was carried out that would result in a product size of 1.4 kb using the same forward primer used in 5.15.A and an internal *GES* primer. The results presented in Figure 5.15.B confirm the absence of the *GES* gene in the five transformants. One possible explanation to the lack of positive *Synechocystis* transformants might be attributed to leaky expression of the toxic *GES* gene under the *nrsB* promoter that caused cell death. Since geraniol has already shown to be highly toxic to the cell, even minute amounts of the expressed gene may lead to cell death.

On the other hand, PCR results presented in Figure 5.15.C confirm the isolation of *Synechocystis* strains with the deleted *GES* gene under the *psbA2* promoter that has been integrated at the *psbA2* locus. Traces of the wild type bands also appear in the PCR results demonstrating that the cells have not yet reached the homoplasmic stage. Nevertheless, these results confirm that once the gene product has been inactivated by partial deletion of the gene, integration and detection of GES at the *psbA2* site was possible (Figure 5.16), thus further confirming that the lack of positive transformants obtained from both the pLAH.A2.GES and pLAH.nrsB.GES was due to either strong or leaky expression of *GES*, respectively.

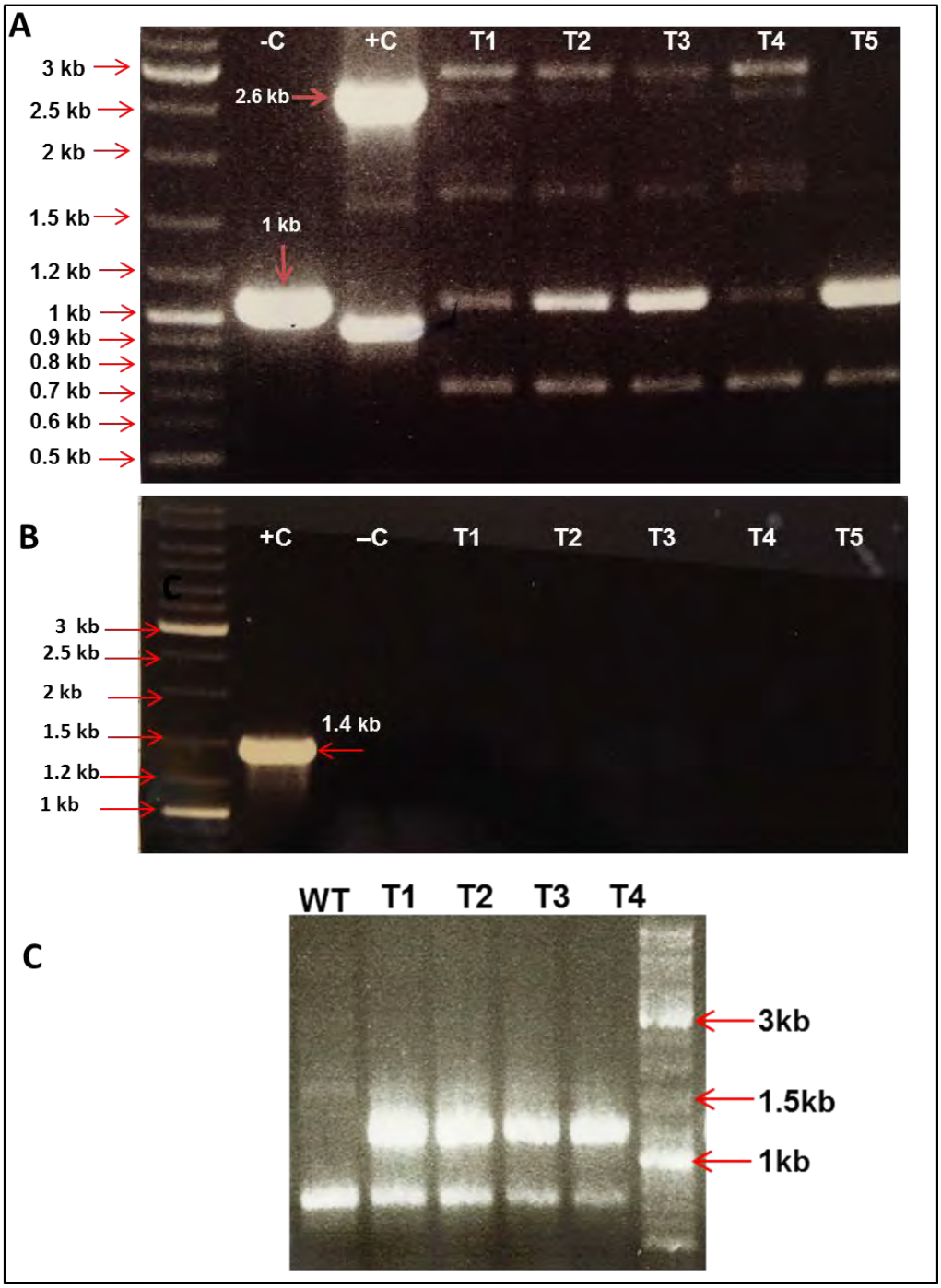


Figure 5.15: PCR results of *Synechocystis* transformed with full and deleted *GES*.

(A) Results show the absence of positive *Synechocystis* transformants for *GES* under the *nrsB* promoter. -C: pLAH.nrsB without the *GES* gene (~ 1 kb). +C: pLAH.nrsB.GES plasmid (~ 2.6 kb). Many non-specific bands appear on the gel. **(B)** PCR analysis of the same five transformants and controls shown in (A) using an internal *GES* primer, **(C)** PCR results showing the isolation of *Synechocystis* strains with the deleted *GES* gene incorporated in the *psbA2* locus. Expected band size ~ 1.2 kb.

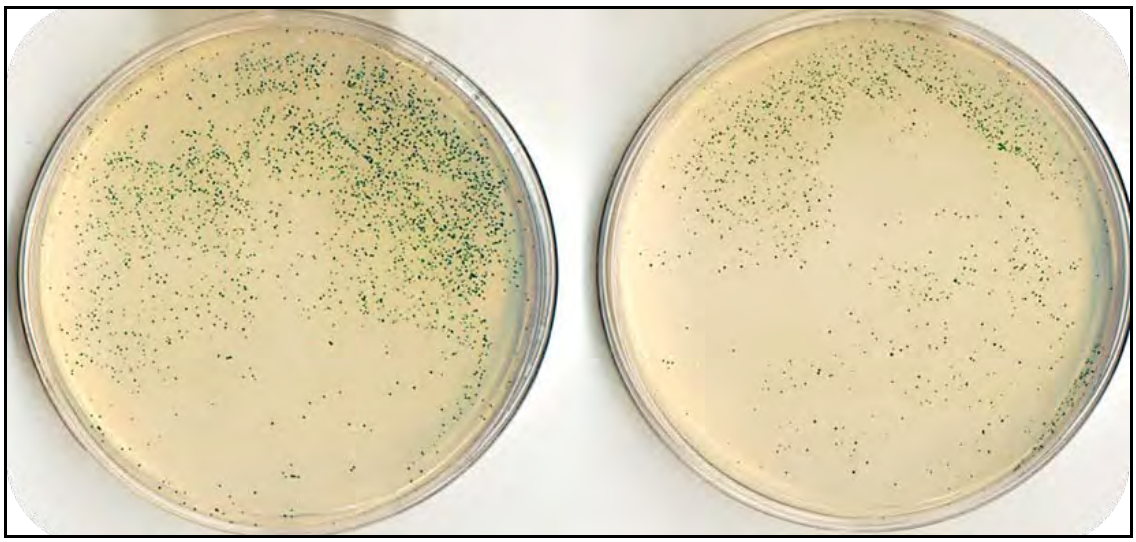


Figure 5.16: *Synechocystis* transformed with a partially-deleted version of *GES*.
BG11 plates supplemented with km (200µg/ml) show the successful isolation of 1000's of transformants recovered from transforming wild type *Synechocystis* with pLAH.A2.GES.D.

5.3.3 Toxicity tests of five selected terpenes *on Synechocystis*

Based on the engine tests conducted by Paul Hellier, geraniol, although an attractive high-value product to produce, was considered not suitable as either a diesel or gasoline replacement fuel. On the other hand, the best candidates as diesel fuels were identified to be geranial and farnesene and the best candidates as gasoline fuels were shown to be citronellene and linalool (Hellier et al., 2013). After exploring the feasibility of metabolically engineering *Synechocystis* for production of these selected terpenes (Figure 5.1) they were selected for further investigation.
This section describes the toxicity tests that were performed on *Synechocystis* using the five selected terpenes.

5.3.3.1 Testing for growth of *Synechocystis* in the presence of five selected terpenes

The toxicity test of the five selected terpenes was set up according to methods described in section 2.7.1. The toxic effect of the tested terpenes can be estimated visually as shown in Figure 5.17. Geraniol and geranial were both extremely toxic and no growth was observed even at the lowest concentration (0.02% v/v). Linalool was less toxic at lower concentrations with the concentration of cells in the flask containing 0.02% linalool close to that of the control (2 $x10^7$ cell/ml and 3 $x10^7$ cell/ml, respectively). Cultures containing farnesene showed an interesting behaviour where the growing cells started to clump together after 16

hours with more clumping observed at higher concentrations (Figure 5.17.C), and by the fourth day the cultures were dead. Citronellene was the only test compound that did not seem to inhibit the growth of *Synechocystis*. Cultures grew well even in the presence of the highest concentration of citronellene (1% v/v). The absence of growth of *Synechocystis* in the flasks was confirmed by an absence of colonies after plating 200 µl of the cultures on solid medium. The viability of the starter culture was confirmed by observed growth in the control flasks.

It can be concluded from the toxicity test that none of the five tested terpenes have the potential to be expressed constitutively (except for small amounts of linalool) in *Synechocystis* due to their inhibitory effect. Citronellene, although showing no observed inhibitory effect on *Synechocystis*, cannot be produced naturally (Figure 5.1) leaving the idea of constitutive expression of the selected terpenes very challenging.

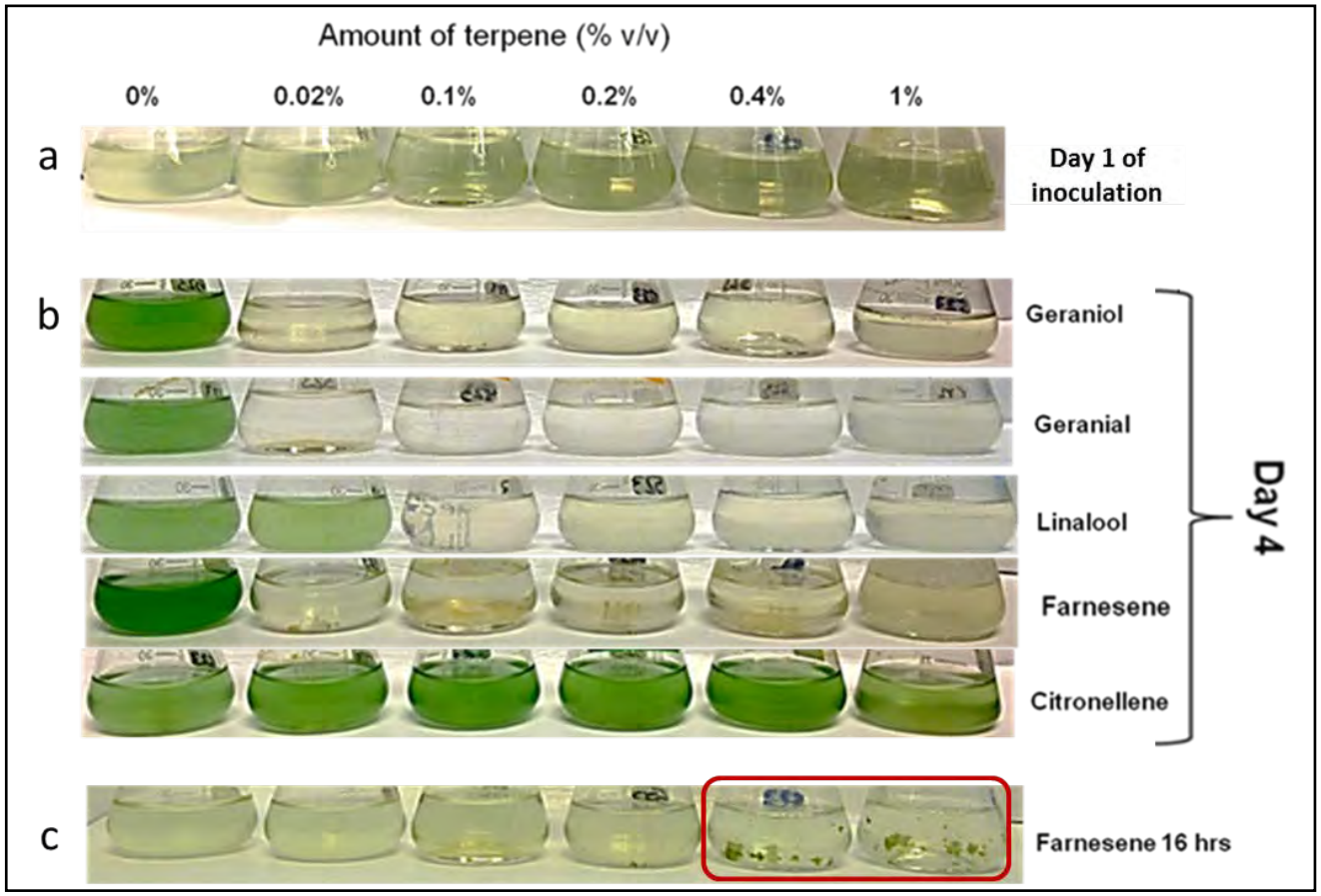


Figure 5.17: Testing for *Synechocystis* growth in the presence of five selected terpenes.

(a) Flasks containing inoculums of *Synechocystis* spiked with increasing amounts of selected terpenes. **(b)** Growth of *Synechocystis* in the presence of increasing amounts of five selected terpenes four days post co-inoculation. **(c)** Clumping of *Synechocystis* cells 16 hours post co-inoculation with higher concentrations of farnesene.

5.3.3.2 Testing levels of tolerance of established cultures of *Synechocystis* to added terpenes

The tolerance of pre-established liquid cultures of *Synechocystis* to geraniol, geranial, linalool, farnesene and citronellene was investigated by monitoring whether addition of the test compound inhibited further growth and led to bleaching of the culture.

As shown in Figure 5.18, geraniol and geranial had the most toxic effects on *Synechocystis* even after 24 hours of addition as shown by the almost complete loss of pigmentation even at a concentration of the 0.02 % (v/v). Linalool was less toxic than geraniol and geranial at lower concentrations while the cyanobacterium was able to tolerate the highest concentration (1 %) of both farnesene and citronellene, suggesting that controlled biosynthesis of these two compounds in genetically engineered strains should not prove toxic, at least at this concentration.

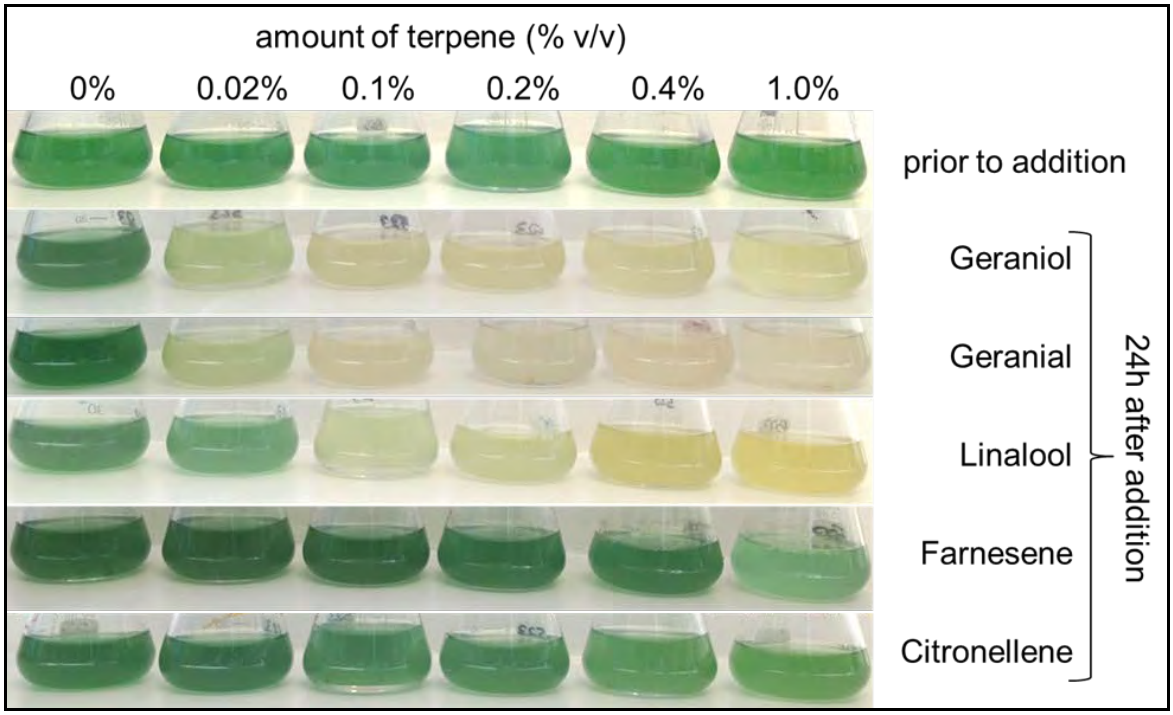


Figure 5.18: The tolerance of *Synechocystis* to increasing amounts of terpenes (24 hours).

Figure 5.19 shows the same experiment after four days. It is clear that the cultures are unable to recover from the addition of geraniol or geranial, even at the lowest concentration. The death of the cells was confirmed by plating aliquots from the 0.02 % geraniol flask onto solid medium, and observing that no colonies appeared following incubation. Whilst reduced growth is observed for 0.02 %

linolool and farnesene, higher concentrations appear toxic despite the promising observation for farnesene at 24 hours (Figure 5.18). The least toxic compound is citronellene and healthy growth of *Synechocystis* is seen even after four days exposure to the highest concentration (1%).

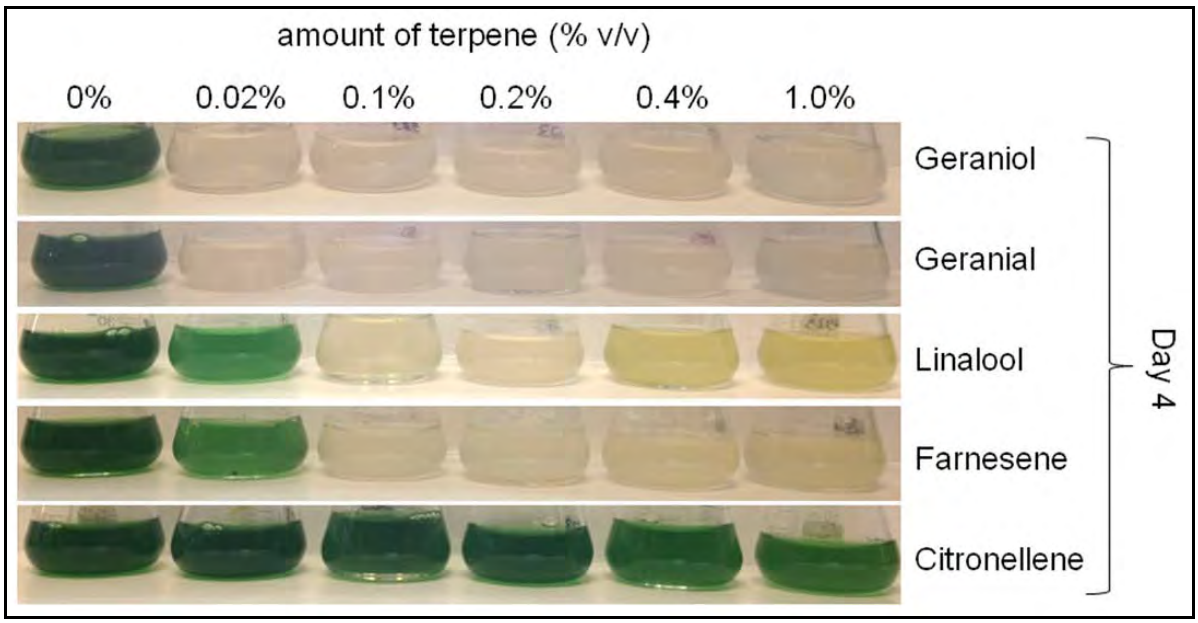


Figure 5.19: The tolerance of *Synechocystis* to increasing amounts of teperens (four days post addition).

5.3.3.3 Effect of a two-phase culture system on tolerance levels

In this experiment the effect of using a two-phase culture system on tolerance levels of *Synechocystis* to terpenes was examined. Figure 5.20 demonstrates the effect of adding n-dodecane to the culture medium, and compares cultures of *Synechocystis* inoculated with 0 %, 0.02 %, 0.04 %, 0.08 %, 0.1 %, 0.2 % and 1 % v/v farnesene (*i.e.* the most promising terpene) grown in the presence and absence of n-dodecane. The flasks containing n-dodecane showed a marked improvement in survival after four days of co-inoculation as seen in Figure 5.20.A, even at farnesene concentrations of 1 %. Figure 5.20.B shows viable cells from cultures grown in the presence of 1% farnesene supplemented with 5 ml of n-dodecane. It can be concluded that adding this non-toxic organic solvent to the culture medium helps improve the survival of cells by forming a two phase system. Consequently, it might be possible to “milk” the toxic farnesene from engineered cells even at concentration up to 1% v/v farnesene.

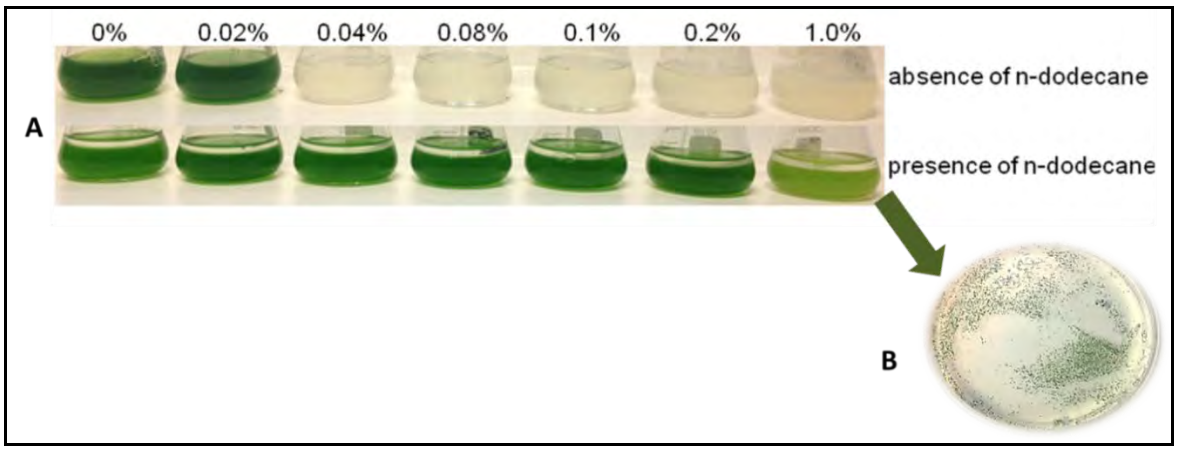


Figure 5.20: Effect of adding n-dodecane on survival of *Synechocystis* in the presence of increasing amounts of farnesene.

(A) Growth of *Synechocystis* in the presence of increasing amounts of farnesene, and in the presence or absence of n-dodecane overlaying the culture medium. **(B)** 200 µl from the culture inoculated with 1.0% farnesene in the presence of n-dodecane plated to check viability of cells on the fourth day.

The effect of n-dodecane on improved survival rates of *Synechocystis* was previously examined for low concentration of geraniol (Figure 5.10). However, the effect of n-dodecane with higher concentration of geraniol was not tested. Therefore, another experiment was set up to investigate the more toxic terpenes namely geraniol, geranial and linalool. In this set up the lowest tested concentration (0.02% v/v test compound) and the highest tested concentration (1.0% v/v test compound) were used in the investigation. The cultures were grown in the presence and absence of n-dodecane for four days. Figure 5.21 shows that n-dodecane helped improve the survival of cultures at the lower concentration of each test compound (Similar results obtained in Figure 5.10) while it did not show any marked improvement in cultures inoculated with the higher concentration of the test compound. Using n-dodecane as a two-phase system to improve tolerance of *Synechocystis* to extremely toxic terpenes to *Synechocystis* (geraniol, geranial and linalool) might not be practical and alternative strategies may need to be adopted to overcome the toxicity issue for higher concentrations of these highly toxic compounds.

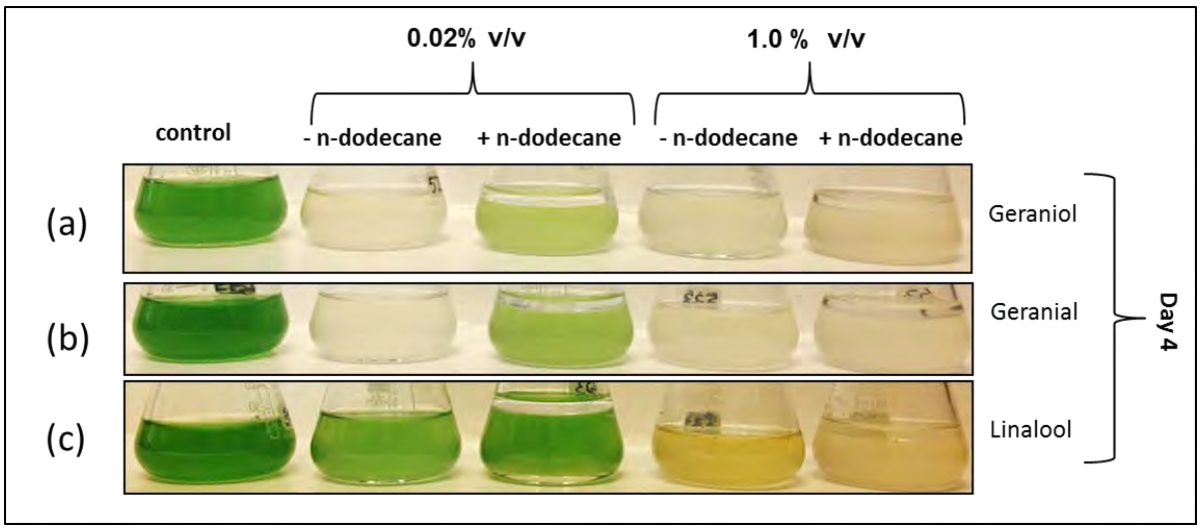


Figure 5.21: Effect of n-dodecane on *Synechocystis* survival in the presence of highly toxic terpenes.

(a) Low (0.02% v/v) and high (1.0% v/v) concentrations of geraniol, (b) low (0.02% v/v) and high (1.0% v/v) concentrations of geranial and (c) low (0.02% v/v) and high (1.0% v/v) concentrations of linalool. In each case, cultures were grown in the presence and absence of n-dodecane where the presence of n-dodecane helped improve tolerance of only low levels of each tested compound.

5.3.4 Expression of a α-farnesene synthase gene in *Synechocystis*

This section describes the synthesis of a synthetic α–farnesene synthase gene and its heterologous expression in *Synechocystis.* It also details attempts to improve expression of the synthase under both the *psbA2* and *nrsB* promoters. It describes GC-MS methods used for detection of farnesene and finally, discusses interesting phenotypic observations of the α-farnesene synthase expressing strains.

5.3.4.1 Synthesis of codon-optimised gene for α–farnesene synthase

In order to obtain high and stable expression rates of α–farnesene synthase in *Synechocystis*, a synthetic, codon optimised, version of the gene was commercially synthesised by GeneArt [10]. The synthesis was based on the mature protein of α–farnesene synthase from *Pyrus communis* (GenBank: AY566286.1) and the codon usage was adapted to the codon bias of *Synechocystis* sp. PCC 6803 genes. The codon optimised gene sequence and its corresponding protein sequence can be found in Appendix 11.

[10] www.geneart.com

5.3.4.2 Construction of plasmids for transformation of *Synechocystis* with the α-farnesene synthase gene

For western blot detection purposes an HA-tag sequence was added to the 3' end of the codon optimised α–farnesene synthase (FS) gene. The addition was done by PCR amplifying the α–farnesene synthase gene from the construct provided from GeneArt using a reverse primer with an attached HA-tag sequence. The primers also included a unique *Nde*I restriction site at the 5' end of the gene that served as the start codon for translation and a unique *Bam*HI restriction site designed immediately downstream of the HA-tag sequence and stop codon. These engineered sites were used for cloning in the corresponding restriction sites of the pLAH.A2 and pLAH.nrsB expression vectors. Details of the primers can be found in Appendix 3.

The PCR product was digested with *Nde*I and *Bam*HI restriction enzymes and cloned into the vectors to create pLAH.A2.FS and pLAH.nrsB.FS. The correct insertion of the α–farnesene synthase gene in the two plasmids was confirmed by digestion using *Nde*I. Empty pLAH.A2 and pLAH.nrsB (lacking the FS gene) were used as negative controls. A band shift of ~1.7 kb from the negative controls confirmed the insertion of the FS gene in both vectors as seen in Figure 5.22. To check for mutations, a representative of each recombinant plasmid was sent for sequencing and results showed that the pLAH.A2.FS carried a synthase that was a 100% match to the codon optimised sequence of the α–farnesene synthase gene (42 bases were not sequenced due to bad signal) while the pLAH.nrsB.FS showed a 99.9% similarity due to the presence of two nucleotide mismatches. However the two mutations were silent and did not result in changes in the amino acid sequence of the synthase (Appendix 17).

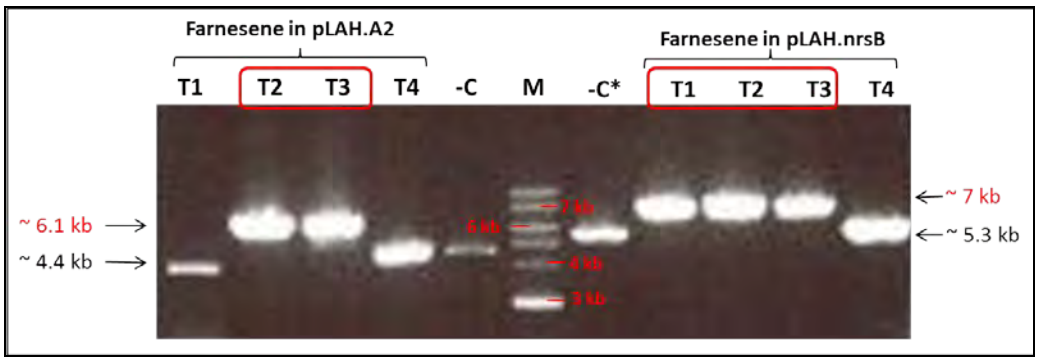


Figure 5.22: Test digest confirming insertion of FS into pLAH.A2 and pLAH.nrsB expression vectors.

Negative controls -C and –C* represent the empty pLAH.A2 (~ 4.4 kb) and the empty pLAH.nrsB (~ 5.3 kb) expression vectors respectively. Correct recombinant plasmids indicated in red rectangles.

5.3.4.3 Transformation of *Synechocystis*

The resulting FS plasmids from section 5.3.4.2 were used for transformation into WT *Synechocystis*, as detailed in section 2.4.2. Transformants were selected on medium supplemented with kanamycin (200 µg/ml). Individual colonies were picked and re-streaked on fresh selective medium and taken through three rounds of selection.

As before, genomic DNA was isolated from eight colonies and analysed by PCR using primers designed to check for correct integration of the FS gene constructs into the *psbA2*. Results from the PCR analysis are shown in Figure 5.23 and confirm the isolation of strains with the α–farnesene synthase gene under the control of the *psbA2* and *nrsB* promoters (hereafter, *Synechocystis* strains 6803.pLAH.A2.FS and 6803.pLAH.nrsB.FS, respectively). Primers used in this PCR gave a product size of ~2.7 kb for positive transformants containing the gene under the *nrsB* promoter and a PCR product size of ~1.9 kb for those containing the gene under the *psbA2* promoter. Primers used can be found in Appendix 3.

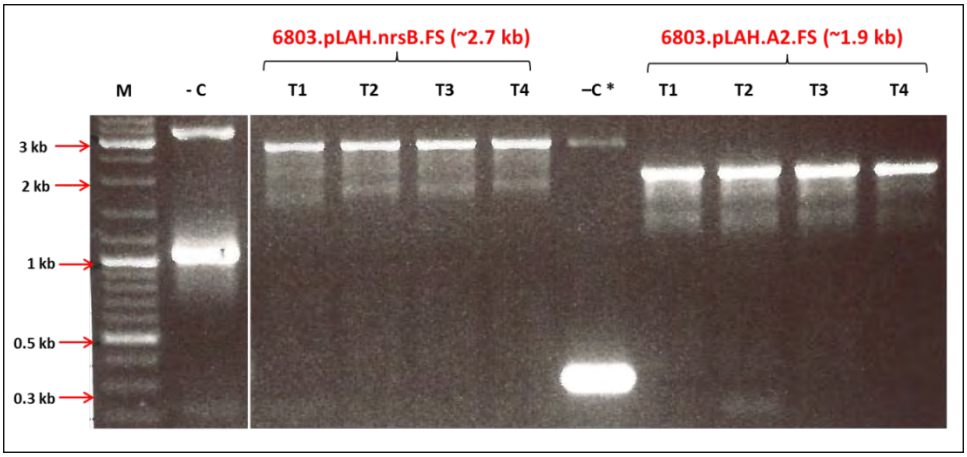


Figure 5.23: *Synechocystis* transformants obtained using FS in pLAH.A2.FS and pLAH.nrsB.FS plasmids.

PCR results confirm the isolation of transformants expressing the α-farnesene synthase gene under the *psbA2* and the *nrsB* promoters. The negative controls -C and –C* represent the empty pLAH.nrsB (~1.0 kb) and the empty pLAH.A2 (~0.25 kb) vectors, respectively. All screened transformants were positive for the FS gene.

5.3.4.4 Western blot analysis of *Synechocystis* transformants

Western blot analysis was used to confirm the constitutive and induced expression of the α–farnesene synthase gene from the transformants obtained in section 5.3.4.3.

Cultures of the 6803.pLAH.A2.FS strain and the 6803.pLAH.nrsB.FS strain were grown to mid log phase. Strains transformed with empty pLAH.A2 and pLAH.nrsB expression vectors were used as negative controls to rule out any non-specific bands that may appear at the expected protein size of 66 kDa. The 6803.pLAH.nrsB.FS and 6803.pLAH.nrsB control strain were then induced with 6.4 µM of $NiCl_2.6H_2O$ for 4.5 hours, after which both cultures were harvested for protein analysis as described in section 2.6. The western blot was set up and the protein detected by ECL using anti-HA antibodies as detailed in section 2.6.3 and 2.6.4 respectively.

Results shown in Figure 5.24 clearly show the presence of the recombinant α-farnesene synthase protein in the *Synechocystis* transformants. The protein was produced under both constitutive and inducible promoter systems where the constitutive expression is evident in Figure 5.24.A indicated by the presence of band at ~66 kDa, and induced synthesis of α-farnesene synthase observed in Figure 5.24.B. In this figure, bands corresponding to α-farnesene synthase were only seen under induced culture conditions and no expression was observed in any of the negative controls confirming the tightly controlled expression of the gene under the *nrsB* promoter.

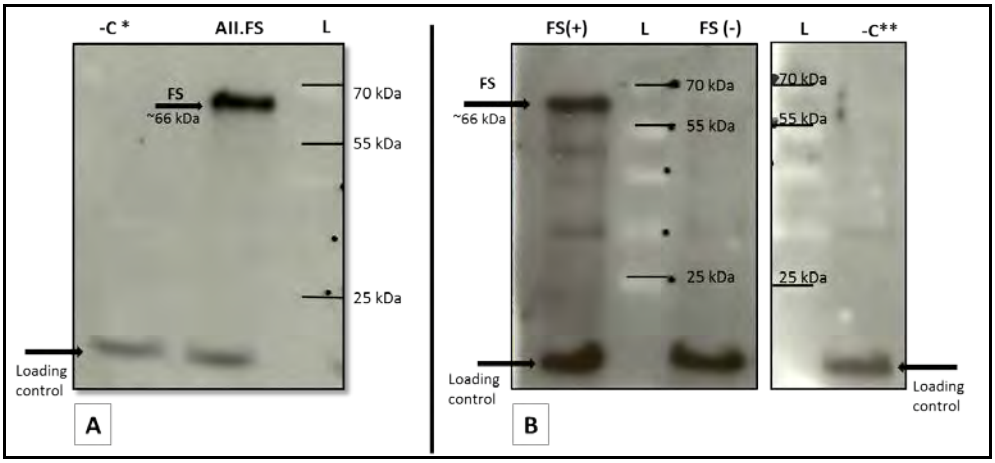


Figure 5.24: Western blot analysis confirms expression of the α-farnesene synthase gene in *Synechocystis* transgenic lines.

(A) Constitutive expression of the gene for α-farnesene synthase (FS) (~ 66 kDa) under the *psbA2* promoter (5 min exposure to X-ray film). (B) Expression under Ni+ induced conditions (15 min exposure). * WT transformed with an empty pLAH.A2, **WT transformed with an empty pLAH.nrsB. The + and – signs correspond to induced and un-induced cultures, respectively.

5.3.5 Optimising α-farnesene synthase expression

5.3.5.1 α-Farnesene synthase expression under the *psbA2* promoter

To test the efficiency of the *psbA2* promoter on the expression of the FS gene, experiments were conducted where cultures of *Synechocystis* transformants were subjected to growth under different light conditions since the *psbA2* promoter is known to be light induced and regulated (Lindberg et al., 2010). To test expression under low and high light, two independent transformants expressing the FS gene under the *psbA2* promoter were selected for analysis. Each of the transformants was grown under two light conditions, a low light intensity of 20 µmol/m^2/s and a high light intensity of 200 µmol/m^2/s. The cultures were grown at 25^{o}C to an OD_{750} = 0.8-1.0 before being harvested for western blot analysis using anti-HA primary antibodies. Details of antibody concentrations and ECL detection methods can be found in Table 2.3.

Figure 5.25 appears to show that accumulation of α-farnesene synthase (~66 kDa) in cultures grown under high light intensities is higher as compared to the same cultures grown under lower light intensities, indicated by a more intense band. Although these results requires further study to confirm the observation, the finding is in agreement with work done by (Mohamed and Jansson, 1989, Lindberg et al., 2010) who showed that high light intensities enhance expression of the *psbA2* gene.

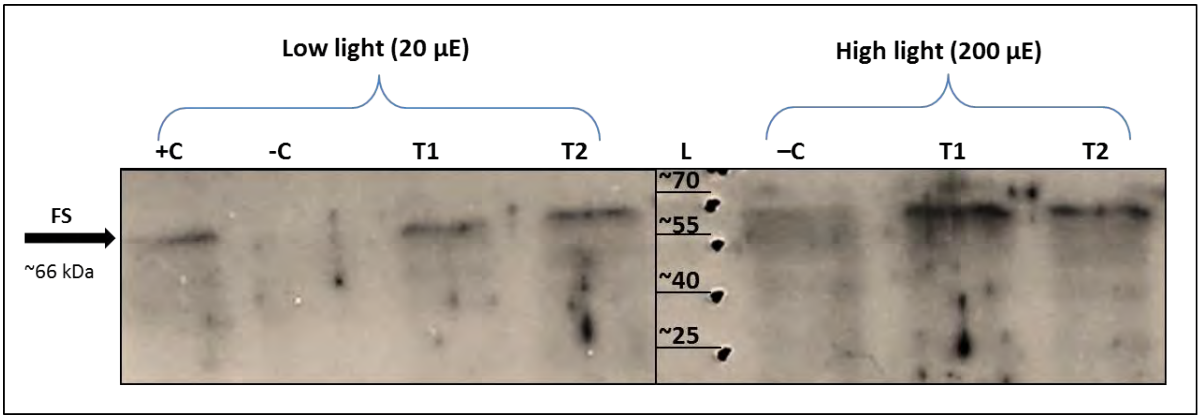


Figure 5.25: Light regulated expression of α-farnesene synthase.

Western blot results showing accumulation of α-farnesene synthase (~ 66 kDa) in *Synechocystis* cultures grown under low light (20 µmol/m^2/s) and high light (200 µmol/m^2/s). No protein is detected with the negative controls. FS: α-farnesene synthase, +C: FS expressing strain, -C: WT 6803 transformed with an empty pLAH.A2 vector.

Work done on optimising the expression of a isoprene synthase (IPS) gene in *Synechosystis* under the *psbA2* promoter showed that the levels of IPS was induced as a function of time under high light, with levels of protein increasing steadily between 0 and 6 hours of post high light incubation (Lindberg, Park et al. 2010). Therefore, the same experiment was performed to investigation the levels of α-farnesene synthase in cultures grown firstly under low light followed by a shift to high light conditions.

For this study, T1 from Figure 5.25 was selected for the analysis with a culture grown under low light intensities (20 µmol/m^2/s) to an OD_{750} of ~0.7 before shifting it to high light conditions (200 µmol/m^2/s). Samples were harvested between t= 0 to t= 6 h after the shift. Figure 5.26 shows that α-farnesene synthase accumulated to a higher levels at t= 0.5h and t= 2h after shifting to high light. Longer incubations of the culture under high light condition appear to result in a reduced signal. Although these results are not consistent with that shown by Lindberg and colleague's. (2010), they still demonstrate the ability of the *psbA2* promoter to regulate the expression of the heterologous expressed FS gene in response to shifting irradiance conditions.

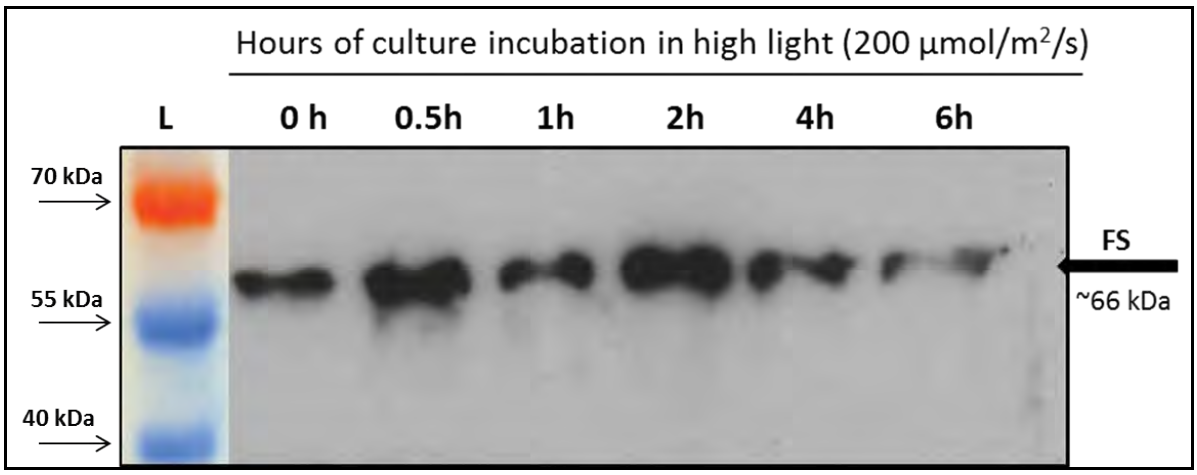


Figure 5.26: Effect of high light treatment on FS expression levels.

Western blot showing higher levels of α-farnesene synthase with cultures initially grown in low light (20 µmol/m^2/s) to the mid log phase then shifted to high light (200 µmol/m^2/s) for 0.5 or 2 hours.

5.3.5.2 α-Farnesene synthase expression under the *nrsB* promoter

Figure 5.24.B showed the induced expression of the α-farnesene synthase gene in cultures exposed to 6.4µM of Ni^{2+} for 4.5 hours. The inducer was added in the cultures at OD_{750} ~ 0.7. The concentration of the inducer and the duration of induction were selected based on published work (García-Domínguez et al.,

2000, López-Maury et al., 2002, Liu and Curtiss, 2009) and were used as reference points for work in this section. However, in order to achieve higher levels of induced FS in *Synechocystis*, it was essential to study different parameters that may affect the expression level. This section describes experiments where the effect of Ni^{2+} concentration; induction duration and stage of growth at time of induction are evaluated.

Effect of Ni^{2+} concentration: To study the effect of Ni^{2+} a transformant expressing α-farnesene synthase gene under the *nrsB* promoter was grown to an OD_{750} of ~0.7. The culture was split into six flasks and $NiCl_2.6H_2O$ added to a final concentration of 0 μM (control), 0.8 μM, 1.6 μM, 3.2 μM, 6.4 μM and 12.8 μM, respectively. The cultures were induced for 4.5 hours before being harvested for western blot analysis as before.

Work done to evaluate the effect of Ni^{2+} concentration on the accumulation of the *nrs* mRNA in *Synechocystis* showed that the level of expression increased with increased Ni^{2+} concentration (García-Domínguez et al., 2000). The study showed that the operon responsible for *nrs* expression was induced at Ni^{2+} concentration above 0.45 μM and showed a steady increase in expression with higher concentrations of Ni^{2+} up to a maximum of 17 μM. It was therefore expected that higher concentration of the α-farnesene synthase protein would be obtained with increased Ni^{2+} concentration. However, Figure 5.27 shows no obvious trend in FS levels when induced with 0.8, 1.6, 3.2, and 6.4 μM, while cultures induced with 12.8 μM appear to show reduced levels possibility as a consequence of less loading indicated by a weaker loading control. As before, no FS was observed in the negative control (un-induced culture) confirming the tightly controlled expression of the synthase under the *nrsB* promoter.

It is well established that essential transition metals such as Ni^{2+} are toxic to cells at above critical concentrations. However, concentrations of 17 μM of $NiSO_4$ have not been shown to be toxic to *Synechocystis* cultures (García-Domínguez et al., 2000) and therefore in this case, it cannot be concluded that the reduced expression was due to a Ni^{2+} toxicity effect. The loading control suggests that the faint band may have been due to unequal loading of the sample or unequal distribution of the antibody in the membrane. However, the same samples were re-loaded on a new gel and similar results were obtained. Further studies with additional controls are required to establish whether >12.8 μM Ni^{2+} does indeed impact on FS levels.

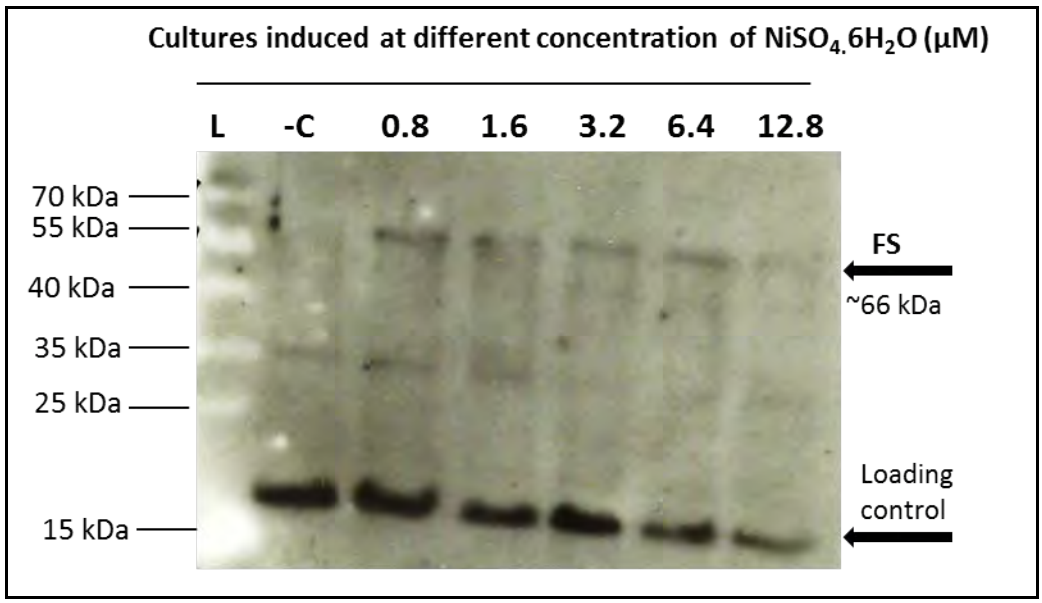


Figure 5.27: Levels of α-farnesene synthase at different concentrations of the inducer.

Similar levels are seen with concentrations of 0.8 µM, 1.6 µM, 3.2 µM and 6.4 µM of the inducer. No FS is seen in the negative control and a very faint band in cultures induced with 12.8 µM of $NiCl_2.6H_2O$. Negative control (-C); WT 6803 transformed with the pLAH.nrsB expression vector.

Effect of induction duration: In order to study the effect of induction duration on the levels of α-farnesene synthase, the same strain used for the Ni^{2+} concentration dependent experiment was grown to an OD_{750} ~ 0.8. The culture was split into eight flasks and 6.4 µM of the inducer was added to each flask with the exception of the negative control that was left un-induced. The cultures were induced for 0, 1.5, 3, 4.5, 6, 7.5, 9 and 24 hours respectively before being harvested for western blot analysis using the anti-HA antibodies.

Figure 5.28 shows that cultures induced for 7.5 hours resulted in maximum levels of α-farnesene synthase followed by the culture induced for 9 hours while no band was observed in cultures induced for 24 hours. The absence of a band in the 24 hour induced culture may suggest that prolonged exposure to the inducer may show detrimental effects on the expression of α-farnesene synthase in *Synechocystis.* The level of α-farnesene synthase for 4.5 hours appears lower than previous results (Figure 5.24.B) where higher expression levels were observed with the same strain induced under the same conditions. This may be due to the physiological state of the strain at the time of the experiment. Nevertheless, it is evident in this study that cultures induced for ~7.5 hours resulted in the highest levels of α-farnesene synthase under the described conditions.

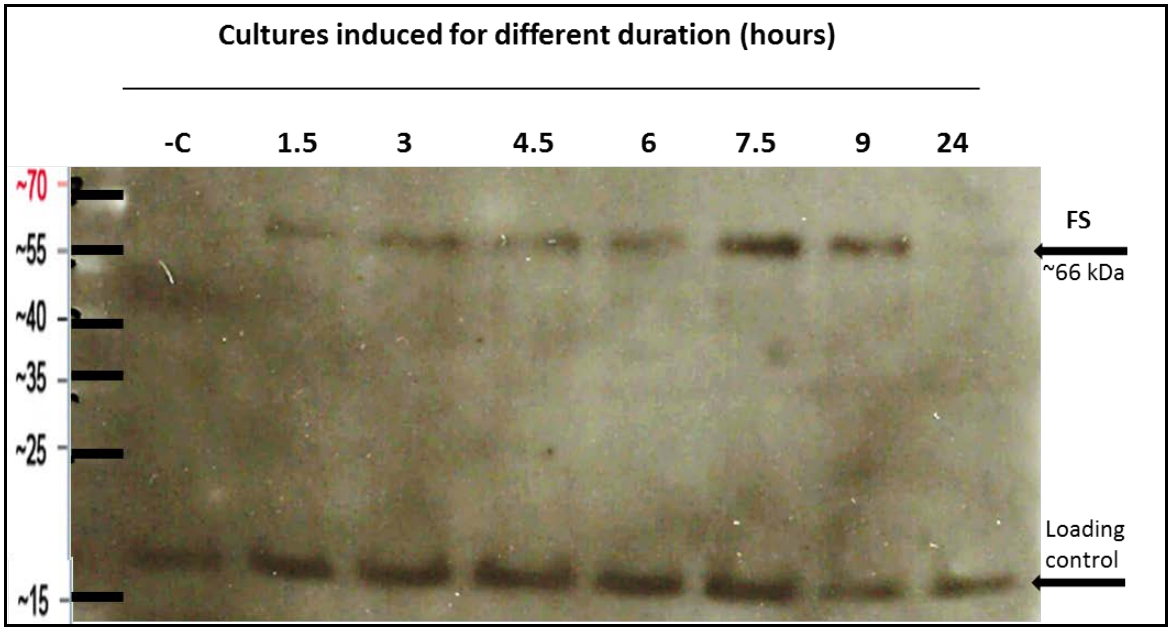


Figure 5.28: Effect of induction duration on the levels of α-farnesene synthase.

Levels of α-farnesene synthase after induction with 6.4 μM $NiSO_{4.}6H_2O$ for different durations (time expressed in hours). Cultures induced for 7.5 hours showed highest expression while the signal disappeared after 24 hours of induction. -C: un-induced culture harboring the FS gene.

Effect of growth stage: Finally, the effect of growth stage at the time of induction on FS levels was investigated. In this experiment, a 150 ml culture was grown to an early log phase of OD_{750} 0.3 after which, 25 ml of the culture was harvested and treated as the negative control (un-induced). The remaining 125 ml culture was split into five flaks and inducer ($NiCl_2.6H_2O$) added to a final concentration of 6.4 μM when cultures reached OD_{750} 0.3, 0.5, 0.7, 0.9 and 1.2, respectively. Each culture was induced for 4.5 hours before being harvested for western blot analysis.

For this particular experiment, FS was detected using the Odyssey®infrared imagining system using Anti HA primary antibodies and Odyssey anti rabbit IgG Dylight800 conjugated secondary antibodies, each incubated with the membrane for two hours.

Figure 5.29.A shows the successful synthesis of FS at different growth stages including the early log phase (OD_{750} ~0.3 – ~0.5), mid log phase (OD_{750} ~0.7), late log phase (OD_{750} ~0.9) and stationary phase (OD_{750} ~1.2). In this experiment, the Odyssey system was used to measure band intensity in order to compare FS levels at different growth stages. (Figure 5.29.B) shows that the highest level was obtained with cultures induced at an OD_{750} ~ 0.5 (band intensity

0.11) while a similar value were obtained when cultures were induced at OD_{750} ~ 0.9 and OD_{750} ~ 1.2 with band intensities of 0.08 and 0.07 respectively. The lowest levels were obtained when cultures were induced at OD_{750} ~ 0.3 and OD_{750} ~ 0.7 (both with a band intensity of 0.03). The low level at OD_{750} ~ 0.7 when detected by Odyssey is contrary to what was obtained in previous experiments where ECL was used to detect FS from the same strain, induced under the same conditions of inducer concentration, inducer duration and growth stage at time of induction (Figure 5.29.C). This might be due to using the ECL for detection as the time of exposure can be adjusted to achieve desirable band intensities. It is worth mentioning that in *nrsB* characterization studies (Garcia-Dominguez, Lopez-Maury et al. 2000, López-Maury, García-Domínguez et al. 2002) the inducer was always added when *Synechosystis* cells reached the mid log phase. To the best of my knowledge, no studies on the effect of growth stage at time of induction on expression levels of the protein has been reported.

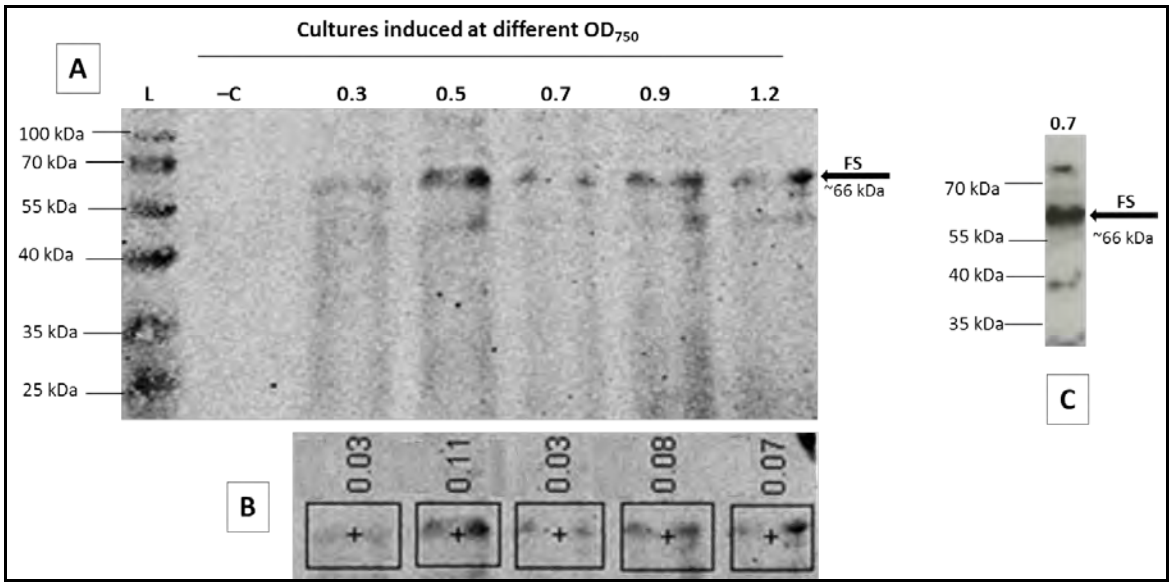


Figure 5.29: Levels of α-farnesene synthase in cultures induced at different OD_{750}.
(A) Induction of cultures at different growth stages. The level of α-farnesene synthase is highest in the culture induced at an OD750 of ~0.5. (B) Band intensity measurements of the α-farnesene synthase. (C) Previous western showing higher levels at OD750 = 0.7.

5.3.6 Farnesene production assays

The next step after the successful expression of the FS gene in *Synechocystis*, under the control of both the *psbA2* and *nrsB* promoters, and optimization of protein levels, was to test cell extracts for active farnesene production.

Supernatant and cell extracts obtained from FS expressing strains under both the *psbA2* and *nrsB* promoters were prepared according to methods detailed in

sections 2.8.1 and section 2.8.2, while samples from the two-phase culture system were harvested according to methods described in section 2.8.3. All the above samples were analysed on a gas chromatography-mass spectrometry (GC-MS) machine according to methods described in section 2.8.4.

5.3.6.1 GC-MS analysis of samples extracted with different solvents

In this section, cell lysates from FS strains that had been extracted using hexane, methanol and ethyl acetate were analysed by GC-MS along with standards represented by hexane spiked with commercially-available farnesene at different concentrations.

The successful detection of the farnesene standard in the three standards (1 ppm, 50 ppm and 100 ppm, respectively) can be seen in Figure 5.30, where a clear peak corresponding to farnesene was obtained at a retention time of 13.95 minutes. Using the selected ion recording (SIR) scanning mode for detection increased the sensitivity of the results through the determination of particular masses present only in farnesene (see section 2.8.4). The standards were used to produce a calibration curve for the quantification of farnesene in the samples extracted from *Synechocystis*. Positive and negative controls, represented by a standard with 1 ppm farnesene in hexane and a methanol extract from un-induced cultures respectively, were included in the analysis. The negative control was included to test the possibility of a false positive (samples contaminated with farnesene) while the positive control was added as a reference for the identification of retention time. Results obtained in Figure 5.31 show that no farnesene could be detected in any of the cell extracts, whether extracted using hexane, methanol or ethyl acetate, while the positive control showed a clear peak at the expected retention time. The absence of farnesene in the samples can be attributed to many possible reasons. Firstly, farnesene might have been produced in very small amounts that were below the detection limit of the GC-MS. Secondly, the biomass used might have not been sufficient to extract detectable amounts of farnesene. Thirdly, the extraction method used (glass beads and solvent extraction) might have not been sufficient to extract small amounts of the product. Furthermore, the amount of the precursors (IPP) and (DMAPP) or the amount of the direct precursor for α-farnesene (FPP) (Figure 5.1) might not have been adequate for the sesquiterpene production. Finally, it is

possible that the recombinant α-farnesene synthase was not active. The following sections investigate the first three possibilities.

GC-MS detection limit: In order to determine the GC-MS detection limit for farnesene, standards of 0.5 ppm 0.25 ppm and 0.125 ppm farnesene in methanol were prepared and run on the GC-MS using the same methods described in section 2.8.4. Figure 5.32 shows the detection of farnesene even in the lowest standard of ~0.1 ppm (100 ppb) confirming that the GC-MS method used was capable of detecting trace amounts of farnesene.

Effect of amount of biomass used: an attempt was made to evaluate the effect of increasing the biomass and concentrating the cultures used for farnesene detection. In this method, the culture size of α-farnesene synthase expressing strain and a negative control (WT *Synechocystis* transformed with empty pLAH.nrsB vector) were increased to 500 ml (20 folds). The cultures were grown to an OD_{750} ~ 1.0 before being induced with 6.4 μM of $NiCl_2.6H_2O$ for 4.5 hours. Pelleted cells from both cultures were analysed for farnesene production. For extraction, in order to concentrate the samples, only 2 ml methanol was added to the pellets along with 500 μl of acid washed glass beads and extraction was carried out as described in section 2.8.2 before being transferred to GC vials for analysis. Unfortunately, no farnesene was detected using this method suggesting that biomass was not a limiting factor in terms of farnesene detection at the conditions tested (Figure 5.33).

Effect of the extraction method: to examine the efficiency of the extraction, which in this case involved vortexing with glass beads in the presence of a solvent, experiments were undertaken using different extraction methods. Sonication for 2 x 15 min in a sonication bath was used as an alternative to glass beads to extract the product using the same three solvents. However, no farnesene was detected in samples prepared with this method (data not shown) suggesting that the absence of farnesene in the samples tested was probably not due to the breakage method.

A further investigation involved the use of different solvents in a “two-phase extraction method” to test extraction of farnesene from the supernatant. For this the non-toxic solvents (decane and n-dodecane) were used. The use of decane, as an extraction solvent has been previously reported (Yoon et al., 2008). This experiment is described in detail in the following section.

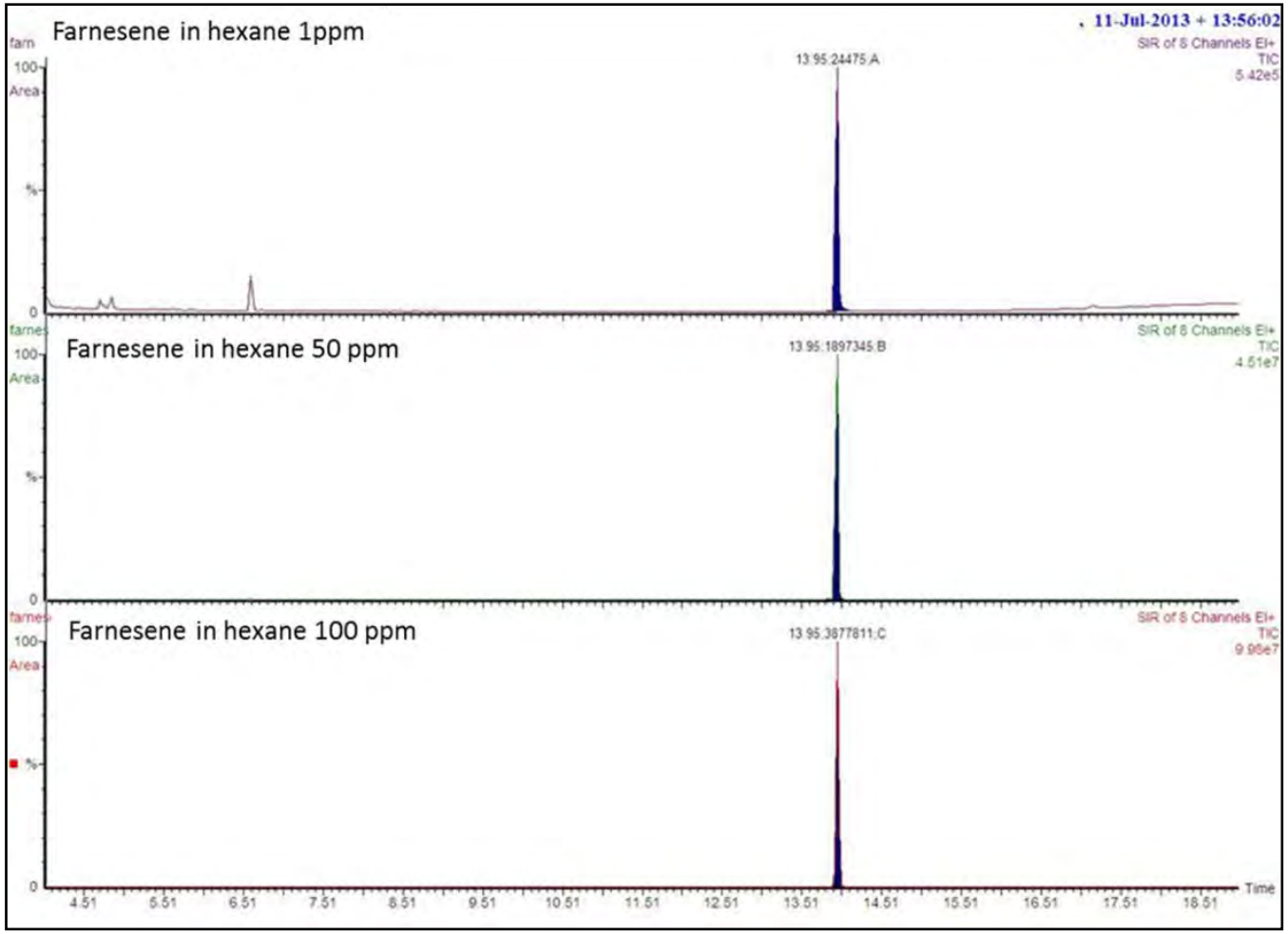


Figure 5.30: Standards of farnesene in hexane at concentrations of 1 ppm, 50 ppm, and 100 ppm.

Peaks corresponding to farnesene obtained at a retention time of ~ 13.95 min.

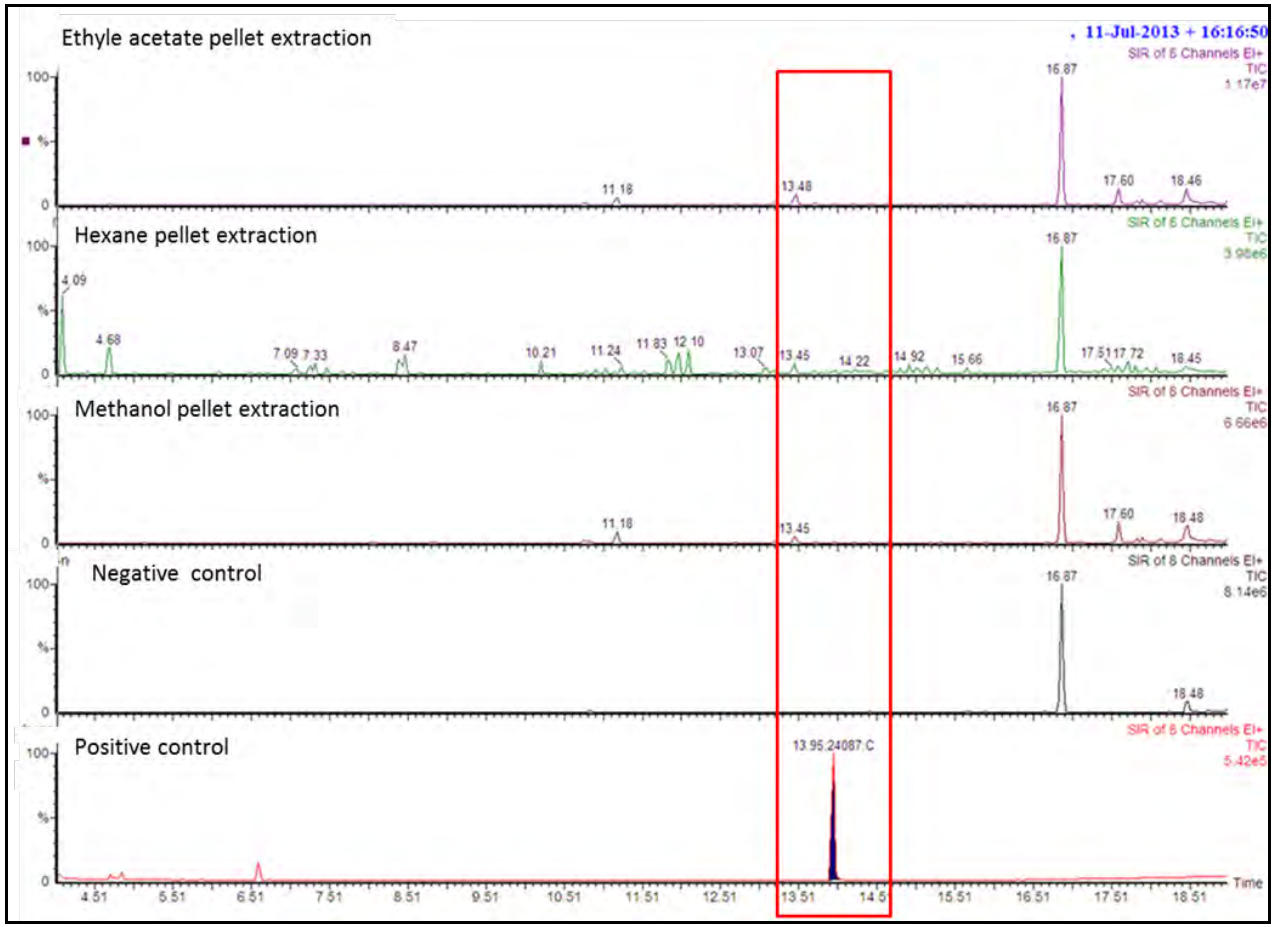


Figure 5.31: GC-MS results for samples containing FS extracted in three different solvents.

GC-MS results showed no detection of farnesene (RT=13.95 min) in pellet extraction with any of the three solvents. The negative control was a methanol pellet extraction from un-induced cultures. The positive control was a standard of farnesene in hexane (1 ppm).

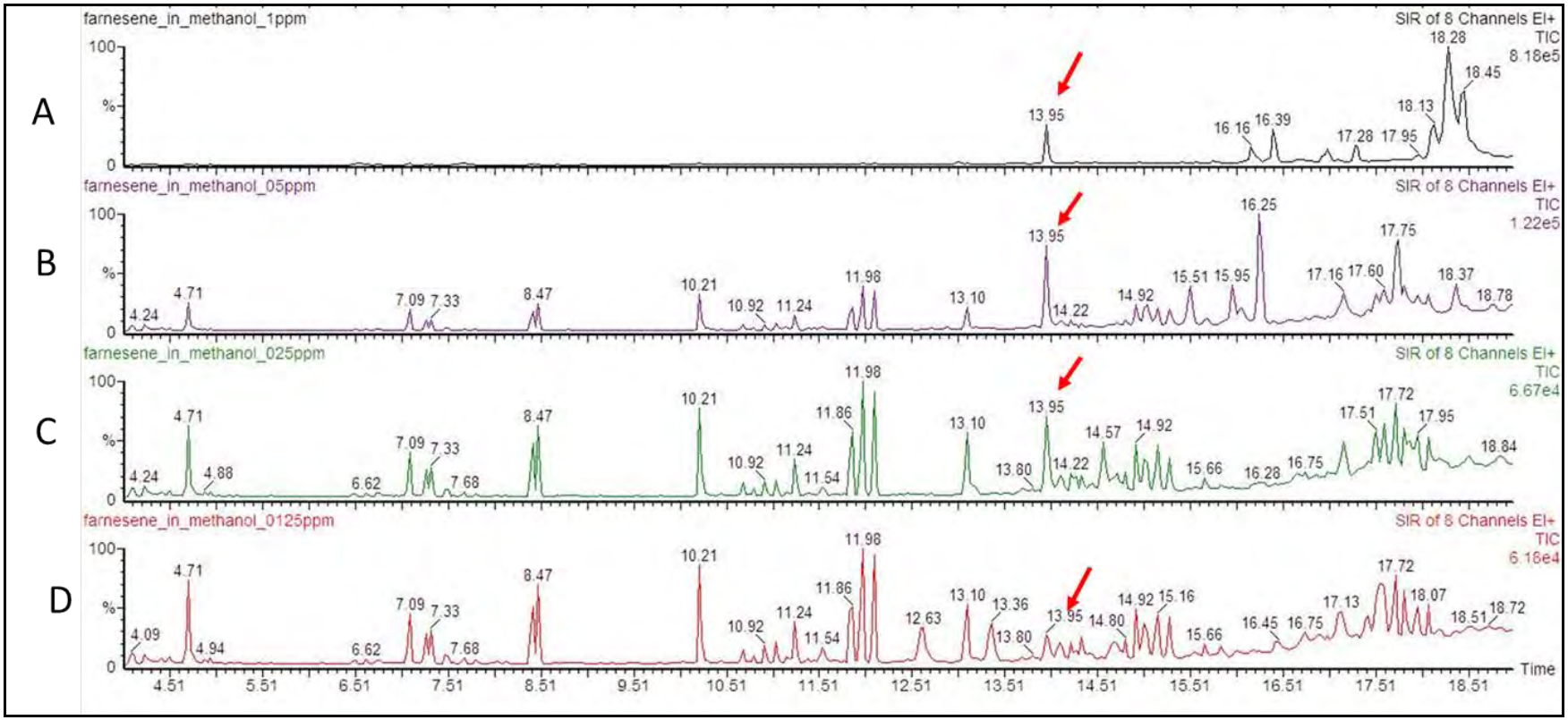


Figure 5.32: GC-MS results for standards of farnesene in methanol

Farnesene detected in methanol cell extracts of 1 ppm (A), 0.5 ppm (B), 0.25 ppm (C) and 0.1 ppm (D).GC-MS can detect farnesene in the lowest prepared standard of ~ 0.1 ppm.

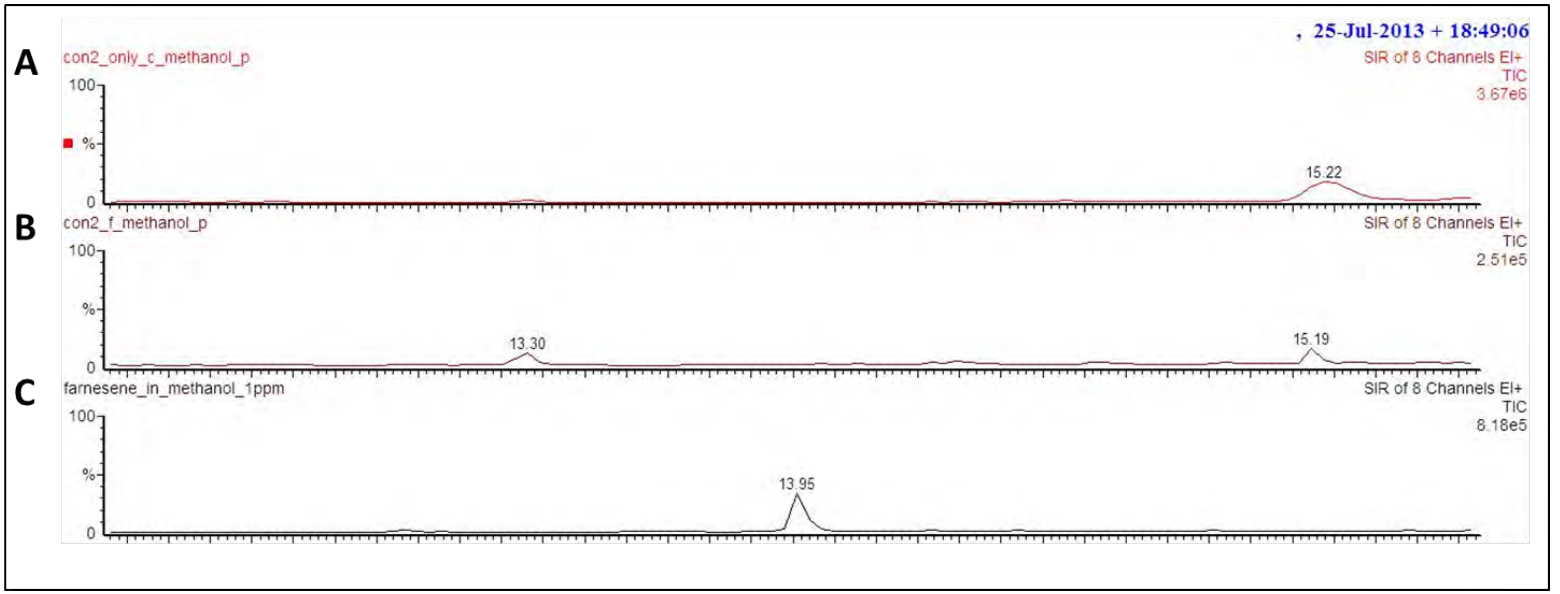


Figure 5.33: Testing the presene of farnesene in samples with increased biomass.

(A) Negative control showing no farnesene. (B) α-farnesene synthase strain with increased biomass showing lack of farnesene in the culture analysed (C) A standard of farnesene in methanol showing farnesene at RT=13.95 (positive control).

5.3.6.2 GC-MS analysis of sample extracts form two-phase extraction system using decane and n-dodecane

In this section, GC-MS results for samples and controls prepared using the two-phase extraction method described in section 2.8.3 are presented. Standards of farnesene in decane or n-dodecane were also analysed.

Two-phase extraction system using decane

Decane was used as a solvent for product extraction and results in Figure 5.34 shows the successful detection of farnesene in standards prepared in decane at concentrations of 1 ppm, 50 ppm and 100 ppm respectively where all three standards showed a clear peak at a retention time of ~13.95, which is similar to that obtained for farnesene in hexane (Figure 5.30). The supernatant and cell extracts from the negative controls showed no peak at RT~13.95 (Figure 5.35) while GC-MS results for samples spiked with 0.02% and 0.2% (v/v) farnesene showed a peak at the expected retention time in both the supernatant and cell extracts. The sample spiked with 0.02% v/v farnesene shown in Figure 5.36 was used as a representative for the results. The detection of farnesene in supernatant extracts of spiked samples confirms the reliability of the extraction method. The fact that farnesene was also detected in the pellet extract of cells gives us an insight on the nature of the interaction between farnesene and *Synechocystis* cells. The exogenously added farnesene in the culture medium probably adheres to the surface of the cells thus resulting in a signal for farnesene on the GC-MS. However, GC-MS results of endogenous α-farnesene synthase from the transgenic strains showed no farnesene in either the supernatant or cell extracts. This can be seen in Figure 5.37.

Similar results for standards, controls and samples prepared with n-dodecane were obtained and GC-MS results can be referred to in Appendix 12 (A, B & C respectively).

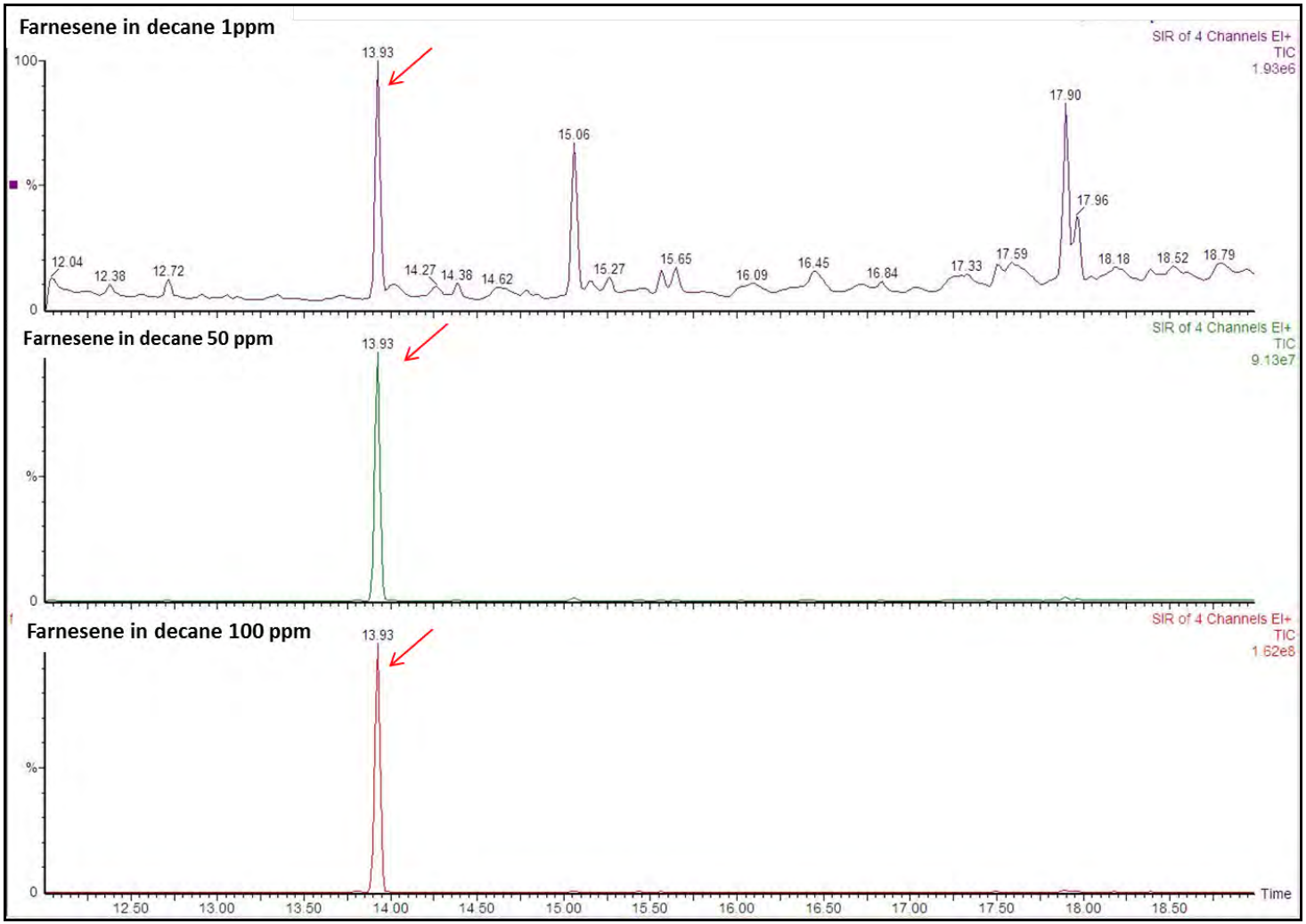


Figure 5.34: GC-MS results of standards of farnesene in decane.

Peaks corresponding to farnesene (RT=13.93) appear in standards of farnesene prepared in decane at concentrations of 1 ppm, 50 ppm, and 100 ppm respectively.

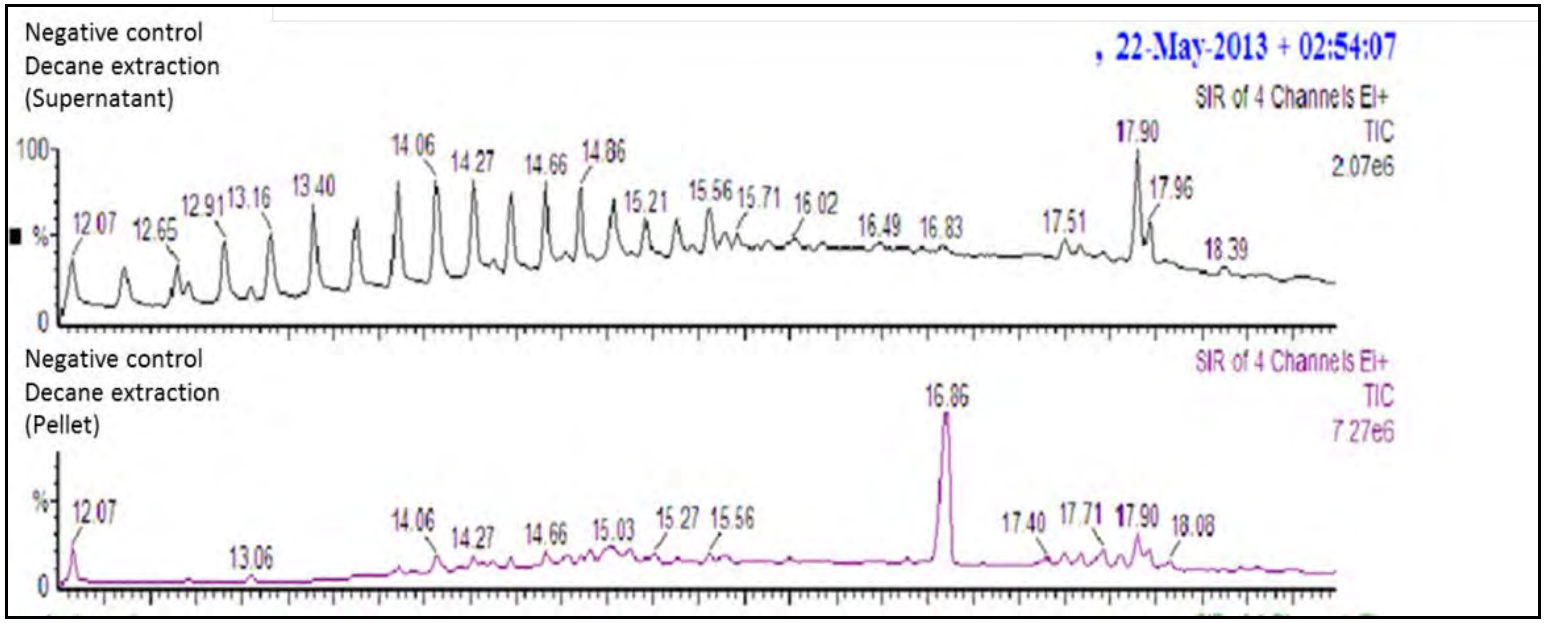


Figure 5.35: GC-MS results of negative controls prepared in decane.

No farnesene (RT~ 13.95) in the negative control samples extracted in decane in both supernatant and pellet fractions. Negative controls are WT 6803 transformed with an empty pLAH.A2 vector.

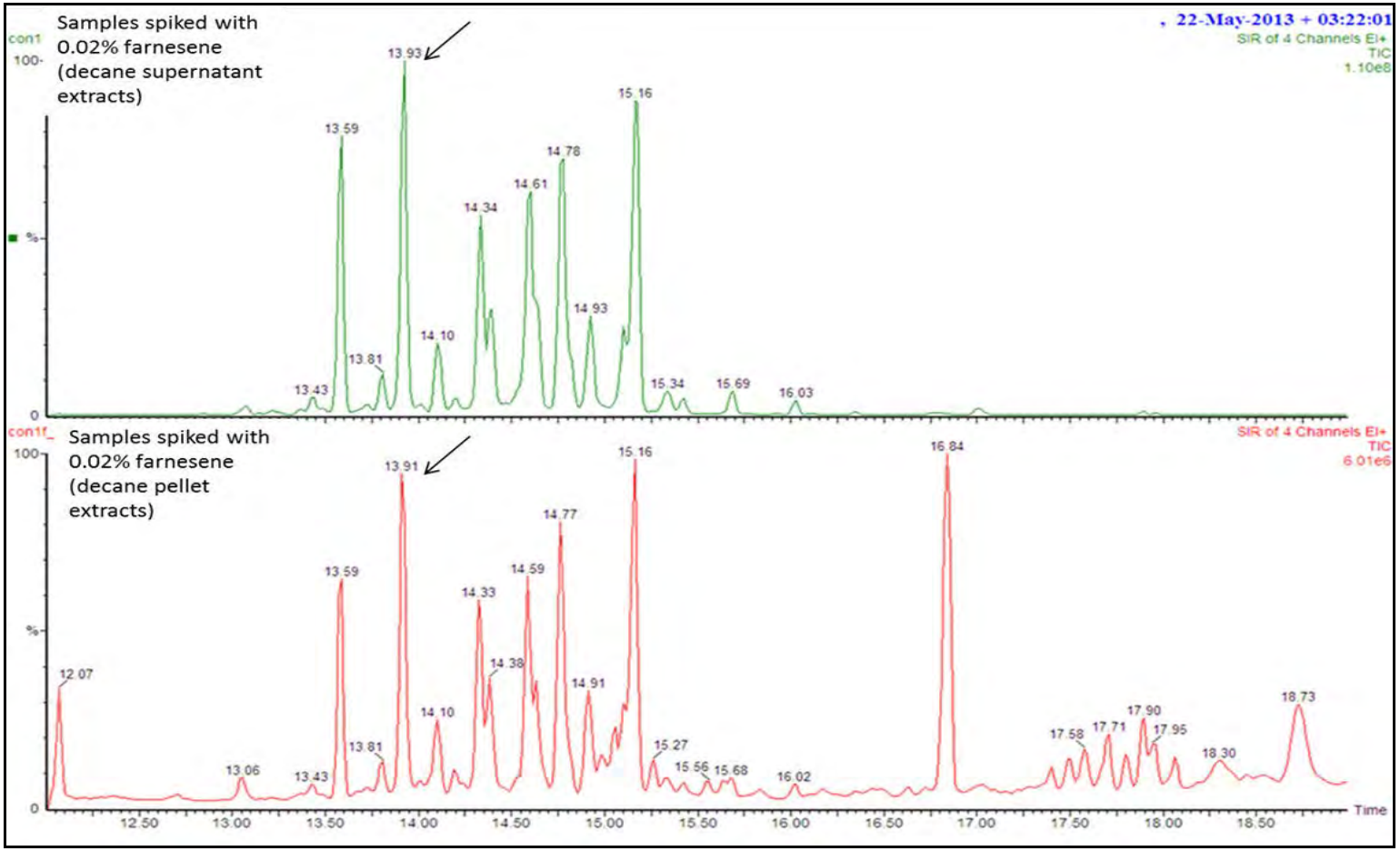


Figure 5.36: GC-MS results of samples spiked with farnesene and extracted in decane.

Farnesene detected by GC-MS from samples spiked with 0.02% farnesene using decane extraction of both supernatant and pellet extracts. The retention time was ~ 13.91 min.

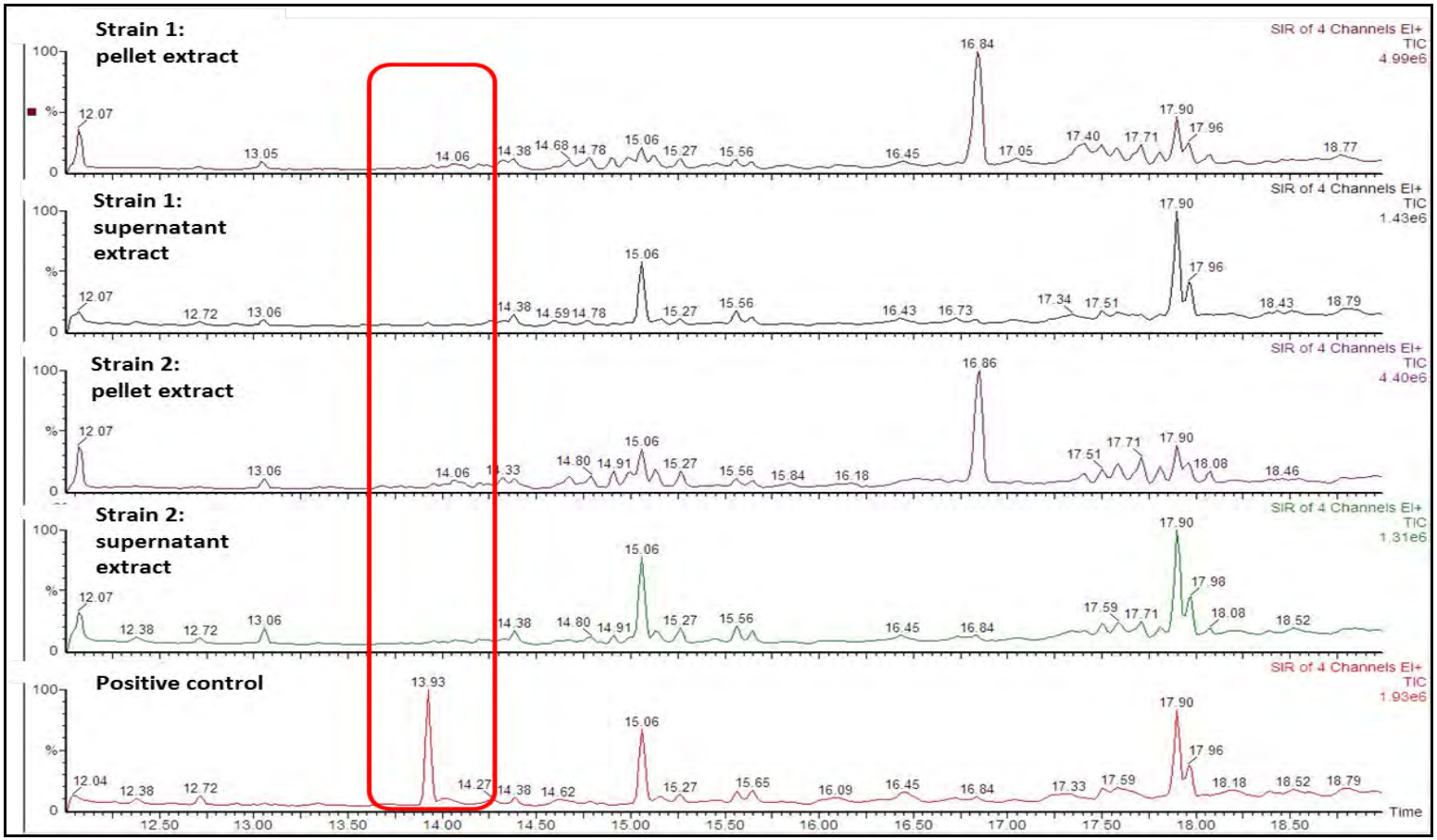


Figure 5.37: GC-MS results of decane supernatant and cell extracts of *Synechocystis* strains.

GC-MS results are of the decane supernatant and cell extracts of *Synechocystis* strains 1 and 2. These represent *Synechocystis* strains expressing the α-farnesene synthase gene under the *psbA2* and *nrsB* promoter, respectively. No peak corresponding to α-farnesene is observed at RT~ 13.95 min for any of the samples. The positive control is a 1 ppm standard of farnesene in decane showing a clear peak at the expected retention time.

5.3.7 Growth analysis & phenotype observation

Despite the inability to detect α-farnesene by GC-MS in strains that were shown to express the α-farnesene synthase gene under the *psbA2* and *nrsB* promoters, further analysis was conducted. In this section, results from growth analysis and phenotype observation are reported for the strain expressing the gene under the inducible promoter. A twin compartment photobioreactor (the Algem 1.0 (Algenuity, UK)) was used for cultivation and growth curve construction as described in section 2.9.

The Algem monitors cell growth based on absorption at a wavelength of 525 nm. Initially, the absorbance of the log phase of the strain was determined under un-induced conditions until it reached stationary phase (OD_{525} ~2.1). The log phase of the strain at an OD_{525} was determined to be between ~0.6–1.6 (Figure 5.38).

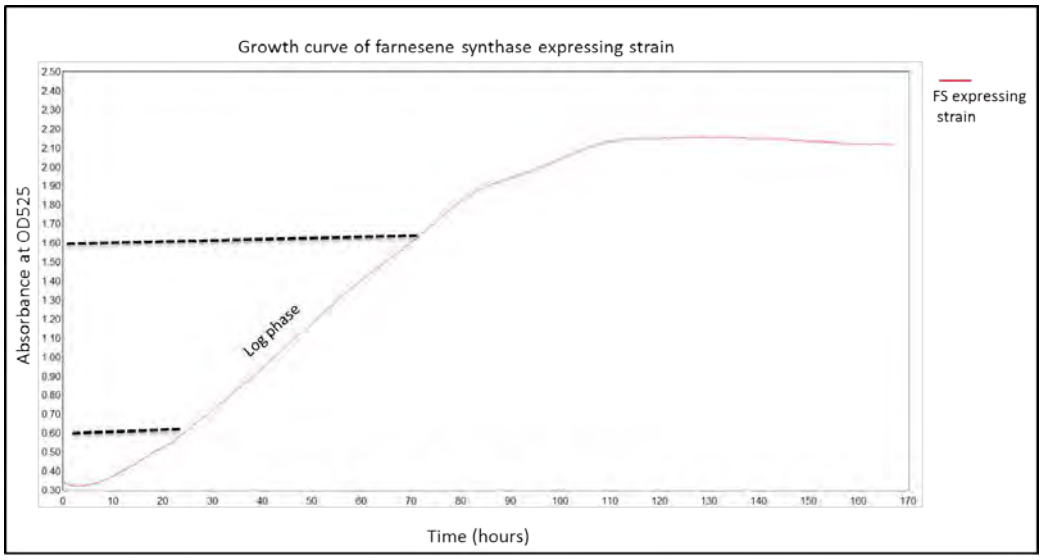


Figure 5.38: Determining the log phase of α-farnesene synthase expressing strain at OD_{525}.

After determining the log phase of the strain of interest, the α-farnesene synthase strain along with a negative control (wild type *Synechocystis* transformed with pLAH.nrsB) were grown from similar OD_{525} starter cultures. The two strains were grown to the early log phase (OD_{525} ~ 0.55) before being induced with 6.4 µM of $NiCl_2.6H_2O$. The strains were allowed to grow to the stationary phase under conditions described in section 2.9.

The growth of the two strains were analysed both before and after Ni^{2+} induction to determine whether there was a significant difference in the growth pattern between the two strains post induction (*i.e.* after expressing the FS gene). Figure

5.39.A shows similar growth patterns of the two strains before nickel induction indicated by the steady increase in growth over 48 hours. Figure 5.39.B on the other hand, which shows the growth pattern of the same strains over 11 days post nickel induction, shows a retarded growth of the FS strain as compared to the negative control. Furthermore, the liquid culture of the FS strain appeared less healthy as compared to the negative control with the pigmentation less green as seen in Figure 5.39.C. The fact that the FS strain showed both a reduced growth pattern and an un-healthy appearance post nickel induction suggests that the synthesis of the α-farnesene synthase is indeed active and imposes a metabolic burden on the cell. To confirm the results obtained and to allow more time for growth pre-nickel induction, the experiment was repeated where the cultures were grown under un-induced conditions for three days. Figure 5.40.A shows the FS strain (green) out growing the negative control (red). Clearly, both strains showed marked growth within the three days of culturing prior to induction. After establishing a healthy growth for both strains, the inducer was added to both cultures when they reached an OD_{525} of ~ 0.9. An interesting phenomenon was observed two days post nickel induction where the FS strain showed a disturbed growth pattern and around day six the culture showed a sudden drop in the OD while the negative control continued to show a normal and healthy growth pattern (Figure 5.40.B).

In terms of appearance between the two strains, Figure 5.40.C shows the FS strain had a yellowish appearance indicating an un-healthy culture post Ni^{2+} induction while the negative control appears healthy (blue-green). The fact that the negative control maintains a healthy culture color confirms that the nickel at concentration used for induction (6.4 μM) has no detrimental effect on the cells and that the yellowish culture color of the FS strain is in fact a consequence of the expression of the α-farnesene synthase gene.

The reproducibility of the results showing the difference in growth and appearance between the FS strain and the negative control post Ni^{2+} induction is a clear indication that the expression of the FS gene in the cell is adding a metabolic burden to the cell leading to deteriorating growth that ultimately leads to death of the culture. It has been reported that introducing transgenes that provide exogenous functions often result in the reduction of the overall fitness and robustness of the recombinant cell compared to native cells, and this might

be a consequence of intracellular depletion of essential precursors (Zhang and Lynd, 2005, Den Haan et al., 2007).

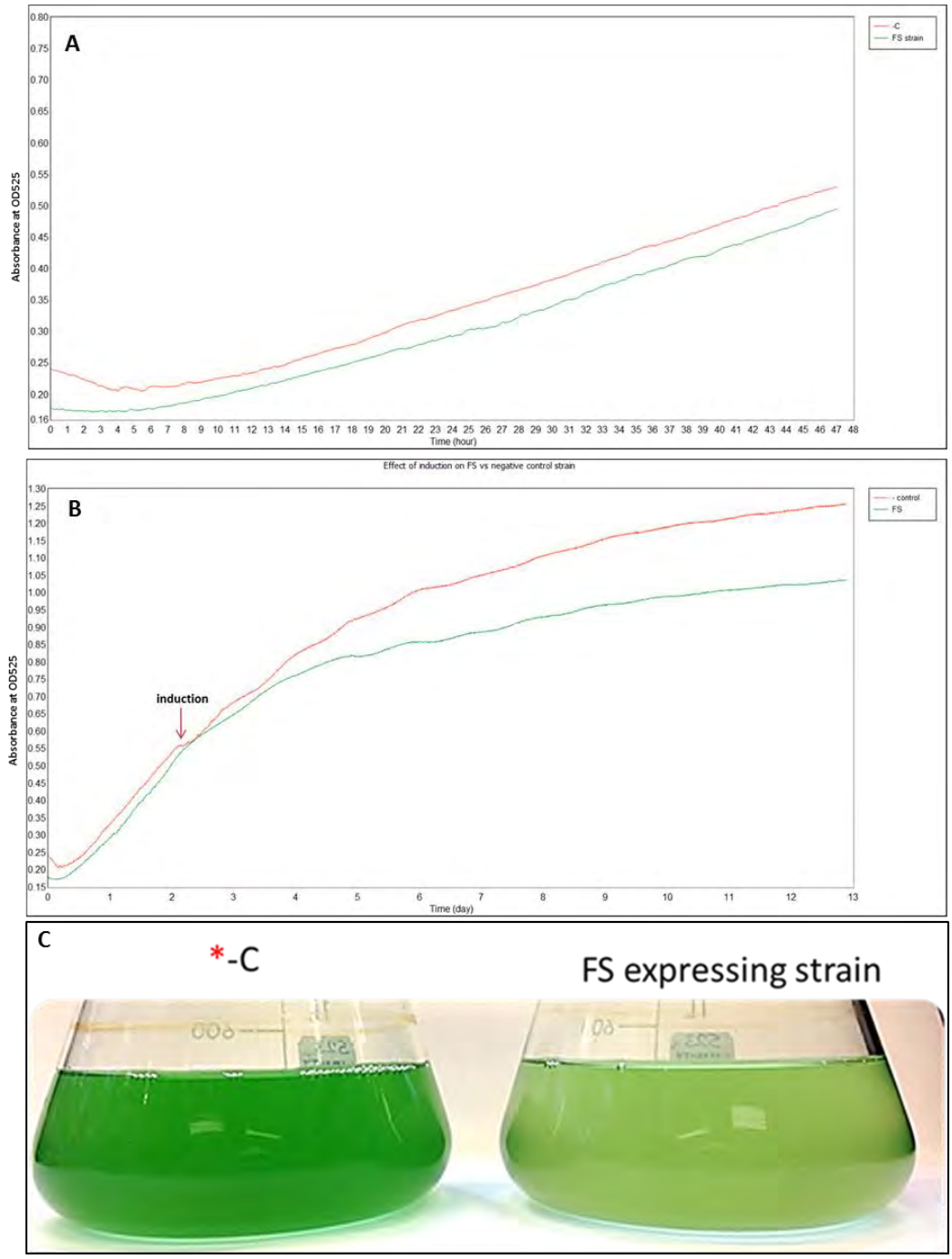


Figure 5.39: Growth of FS expressing strain and negative control pre and post nickel induction.

(A) The growth curve shows similar growth rates for both the α-farnesene synthase strain and the negative control pre-nickel induction, but (B) The strain grows at a slower rate as compared to the negative control post nickel induction. (C) α-Farnesene synthase strain clearly grows less well as compared to the negative control post nickel induction. * WT 6803 transformed with pLAH.nrsB.

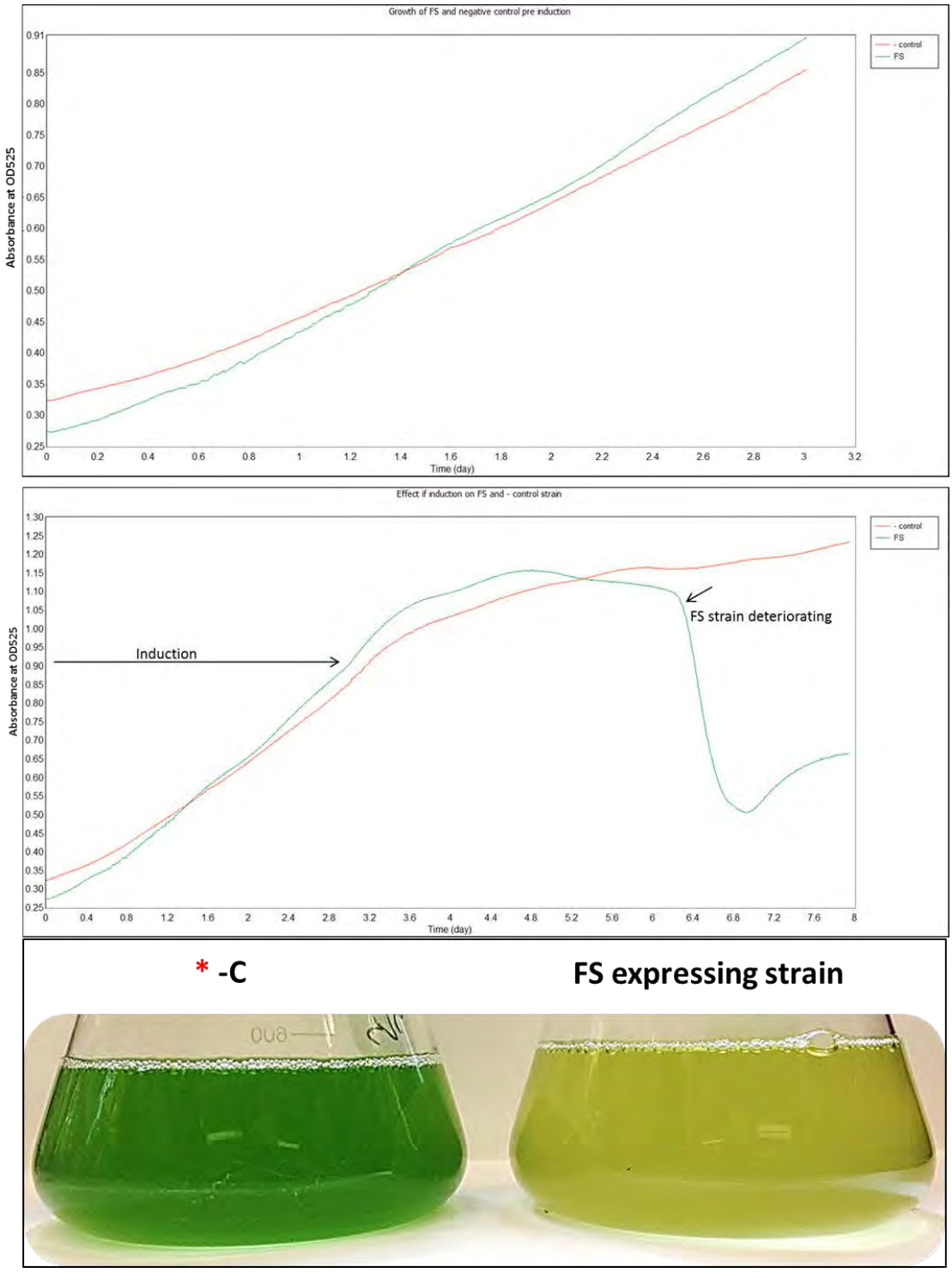


Figure 5.40: Growth of FS expressing strain versus negative control pre and post Ni^{2+} induction.

(A) Growth pattern shows the FS strain (green) out growing the negative control (red). (B) The FS strain showed a sharp and sudden drop about three days post induction while the negative control continues to grow. (C) FS strain looks unhealthy (yellowish) as compared to the negative control (green) post nickel induction.* WT 6803 transformed with an empty pLAH.nrsB.

5.4 Conclusion and future work

Two expression vectors have been successfully developed that allow for the expression of foreign genes in *Synechocystis* under the control of a constitutive (*psbA2*), and an inducible promoter (*nrsB*). Following this success, the pLAH.A2 and the pLAH.nrsB expression vectors have been used by fellow researchers within the group who were able to produce significant quantities of foreign proteins such as limonene synthase (an enzyme for limonene production, a potential biofuel), bacterial antimicrobial proteins, an operon of selectable markers, and human growth hormone (unpublished results). These vectors can be further utilised for the expression of a range of genes of interest in *Synechocystis* with a goal of developing *Synechocystis* as a generic photosynthetic platform for the controlled expression of a wide range of foreign proteins.

The idea of creating an inducible expression vector was primarily to allow for the synthesis of the highly toxic monoterpene, geraniol. However, due to the highly toxic effect of geraniol on *Synechocystis,* we failed to generate transformant lines containing the *GES* gene, suggesting that even basal expression of GES under un-induced conditions is lethal; In order to achieve monoterpene production in microorganisms, it is important to solve this core issue (Brennan et al., 2012). This is because metabolically engineered organisms are forced to produce compounds that may lead to growth autoinhibition (Angermayr et al., 2009). It has been reported that concentration of 0.05% (v/v) geraniol was sufficient to inhibit cell growth of *E. coli* (Dunlop et al., 2011). Using the “two-phase culture system” to capture geraniol did help improve tolerance of *Synechocystis* cultures to small amounts of geraniol (0.02% v/v) while it failed to show improved tolerance with cultures grown in the presence of high concentrations of geraniol (1%). However, the toxicity tests conducted are based on the exogenous addition of geraniol into the culture medium while tolerance level may be considerably different when geraniol is produced within the cells. Furthermore, an alternative method to reduce monoterpenes toxicity could be by expressing an efflux pump to export the toxic compounds out of the cell upon production while using a two-phase culture system for continues harvesting. The potential of expressing a potential efflux pump in *Synechocystis* is the topic of chapter 6.

Farnesene is a potential fuel (Lee et al., 2008), that showed good engine performance when tested in a diesel engine together with promising results in toxicity tests on *Synechocystis* where it was the least toxic even at high concentrations (1%). Farnesene did eventually prove toxic to *Synechocystis* after four days of incubation, however using a two-phase culture system significantly reduced the toxic effects of high concentrations of farnesene. In terms of metabolic engineering work, the farnesene synthase gene was successfully expressed in *Synechocystis* under both constitutive and inducible conditions. Optimization experiments showed that improved levels of α-farnesene synthase expression can be achieved under the *psbA2* promoter when cultures are initially grown under low light followed by a two hour exposure to high light intensities. On the other hand, attempts were made to maximize the induced accumulation of α-farnesene synthase and results showed that an induction duration of ~7 hours induced at an D_{750} ~ 0.5 produced the best results while the inducer concentration did not seem to make much difference to the level of the enzyme. However, more experiments need to be done where the amount of the protein loaded is quantified and the concentration of protein produced is measured by estimation with a protein standard in order to make more accurate conclusions.

Despite that GC-MS analysis was able to detect farnesene at concentrations as low as 0.1 ppm in the standards; no signal was detected in samples following any of the different extraction methods tested. Furthermore, the detection of farnesene in cultures spiked with farnesene was an indication of a successful extraction method. Nevertheless, a limitation of testing extraction using spiked samples relies in the fact that the compound is only extracted from the culture medium with no indication on the efficiency of extraction from within the cell. Testing the extraction of a native protein would provide a better control to test extraction of compounds from within the cell.

A possible explanation for the lack of detectable farnesene could be due to the limited amount of substrate available for the enzyme. Such scenario has also been suggested by several researchers (Oswald et al., 2007, Glick, 1995, Goff and Goldberg, 1985) who suggested that terpene synthesis is limited by the availability of IPP and DMAPP substrates. Furthermore, Martin et al. (2003), reported low levels of the terpene carvone in engineered *E. coli* strains was also attributed to the limited availability of IPP and DMAPP substrates. In order to test

this hypothesis, experiments would need to be undertaken where the IPP and DMAPP substrates and/or the direct precursors to farnesene such as FPP (Figure 5.1) are added exogenously into the induced culture followed by testing for farnesene production over time. Moreover, the fact that only the FS expressing strain showed a detrimental growth rate post nickel induction, compared to strains lacking the FS gene (Figure 5.40), is an indication of the functional expression of the α-farnesene synthase. However, this conclusion needs to be further verified by conducting an enzyme activity assay.

High levels of farnesene exceeding 14 g/L have been achieved in engineered *E. coli* (Rude and Schirmer, 2009) by introducing the mevalonate pathway from yeast in the engineered strain, thereby providing a second route to IPP/DMAPP synthesis and increasing the pool of these precursors. If the activity of the recombinant α-farnesene synthase is confirmed, then introduction of the mevalonate pathway would be an attractive next step.

Chapter 6: The expression of AlkL in *Synechocystis* to facilitate substrate delivery and product removal of hydrocarbons

6.0 Introduction

Many efforts have recently been invested in finding replacements for fossil fuels utilizing photosynthetic organisms that use solar energy to convert carbon dioxide into “drop-in” compatible fuels in a single biological system (Atsumi et al., 2009, Dexter and Fu, 2009, Lindberg et al., 2010, Liu et al., 2011, Wang et al., 2013). As discussed in chapter 5, one of the main challenges in biofuel production is to reduce the overall cost of production of these fuels, a challenging task due to the high cost associated with fuel extraction from the biomass. It is estimated that up to 60% of the cost of algal-based biofuel production is in the extraction of the product from the cell (Molina Grima et al., 2003). Significant research has been done in an effort to reduce the cost of biofuel production in modified microorganisms. For example, the expression of efflux pumps that enhance product recovery by actively exporting the product out of the cell (Dunlop et al., 2011); using two phase extraction systems (Brennan et al., 2012, Hejazi et al., 2004) and the creation of auto-lysing cells (Liu and Curtiss, 2009). In this chapter the goal was to examine whether the introduction of a bacterial transporter protein into the outer membrane of *Synechocystis* would facilitate diffusion of endogenously produced hydrocarbons (*e.g.* alkanes) out of the engineered strain and into the medium.

The ‘bio-pump’ transporter discussed here (AlkL) builds on research conducted by Christopher Grant at UCL (Grant, 2012) and also published by (Julsing et al., 2012). The AlkL transporter of *Pseudomonas putida* GPo1 (also known as *Pseudomonas oleovorans* GPo1) appears to be involved in the transport of alkanes (van Beilen et al., 2001, Hong et al., 2006, Hearn et al., 2009, van Beilen et al., 1992). When expressed in an *E. coli* strain engineered for alkane production AlkL was shown to enable over 120 fold improvement of uptake of C_{12}-C_{16} alkanes (Grant, 2012). Further work by Grant and colleagues also showed that AlkL could lead to up to two fold higher yields of C_{13}-C_{16} alkanes produced by an *E. coli* strain engineered to produce alkanes. Thus, AlkL appears to serve as a channel to facilitate both influx and efflux of hydrocarbons, depending on the concentration gradient across the membrane. Given that cyanobacteria have a similar Gram-negative arrangement of inner and outer membranes as *E. coli* and *Pseudomonas* (Hoiczyk and Hansel, 2000), then it should be possible to similarly introduce AlkL into *Synechocystis* to facilitate transfer of hydrocarbon products across the cell barrier.

In the present work attempts were made to express the *alkL* gene under different promoters in *Synechocystis,* which is known to naturally produce C_{17} alkanes (Schirmer et al., 2010), and examine whether the *alkL* gene product (i) could be functionally expressed and regulated in *Synechocystis* (ii) would play a similar role in the uptake and increase in intracellular levels of alkanes as that demonstrated by Christopher Grant in the engineered *E. coli* strain and (iii) show an improved hydrocarbon yield (iiii) examine its possible role in hydrocarbon extraction.

GC analysis and data processing presented in this chapter were undertaken in collaboration with Dr Grant (Department of Biochemical Engineering, UCL).

6.1 Literature review

6.1.1. Alkanes

Alkanes are saturated hydrocarbons with the general chemical formula C_nH_{2n+2}. They are found throughout nature as they are naturally produced by a diverse range of microorganisms (Schirmer et al., 2010, Hu et al., 2013). Typically, alkanes with carbon chain lengths of C_4-C_{23} have high energy densities, are highly hydrophobic and are compatible with the existing liquid fuel infrastructure since they represent the major constituents of gasoline, diesel, and jet fuel (Schirmer et al., 2010, Peralta-Yahya et al., 2012). Alkanes can therefore be considered potential “drop-in” fuel replacements.

Although the production of alkanes in many cyanobacteria have been reported since the 1960’s (Han et al., 1968), the majority of biochemical studies for alkane biosynthesis have focused on eukaryotic systems (Cheesbrough and Kolattukudy, 1984, Dennis and Kolattukudy, 1992). Recently however, Schirmer et al. (2010), identified the alkane biosynthesis pathway in cyanobacteria. The authors described the cyanobacterial genes and enzymes involved in the two-step conversion of intermediates of fatty acid metabolism to the fatty aldehyde by a specific reductase, and then into alkanes and alkenes via a decarbonylation reaction (Figure 6.1). The authors also show how the heterologous expression of the two genes in an *E. coli* host strain (that does not naturally produce fatty aldehydes) led to the successful production of C_{13} to C_{17} alkanes and alkenes. The recent discovery of the alkanes biosynthesis pathway in cyanobacteria opened a new door for exploring fungible hydrocarbon fuel production from these photosynthetic organisms.

Figure removed for copyright reasons

Figure 6.1: Schematic overview of alkane (alkene), fatty acid and main competing metabolic pathways in *Synechocystis*.

Key: enzyme genes are indicated in the pathways. **3-PGA**, glyceraldehyde 3-phosphate; **PYR**, pyruvate; **PHB**, poly-β-hydroxybutyrate; **acly-ACP**, acyl-acyl carrier protein; ***ddh***, 2-hydroxyacid dehydrogenase gene; ***phaA***, polyhydroxyalkanoates-specific beta-ketothiolase gene; ***accBCDA***, a muli-subunit acetyl-CoA carboxylase gene; ***lipA***, lipolytic enzyme gene; **aas**, acyl-ACP synthetase gene; ***aar***, acyl-ACP reductase gene; ***ado***, aldehyde-deformylating oxygenase. Modified from (Wang et al., 2013).

Heptadecane, a C_{17} hydrocarbon, is the most abundant alkane reported in cyanobacteria (Schirmer et al., 2010) and along with heptadecene forms the two major constituents of alka(e)nes in *Synechocystis* sp. PCC6803 accounting for 0.1% of the cyanobacterium's dry cell weight (Tan et al., 2011, Hu et al., 2013). In recent years, many researchers have been interested in enhancing alkane production in *Synechocystis*. For example, Hu et al. (2013), showed a five-fold increase in the intracellular levels of heptadecane in an engineered strain of *Synechocystis*. This was done by increasing the number of copies of the genes *FAR* (also referred to as *aar*) and *FAD* (also referred to as *ado*) that encode for the fatty acyl-ACP reductase and fatty aldehyde decarbonylase enzymes in the native alkane biosynthesis pathway, respectively (Figure 6.1). Moreover, Wang et al. (2013), showed that the overexpression of the same two genes in two different loci of the genome resulted in an 8.3 fold increase of alka(e)nes production. They also demonstrated that redirecting the carbon flux to acyl-ACP can improve alka(e)ne production in cyanobacteria significantly and that the AAS enzyme encoding acyl-ACP synthetase gene is vital for alka(e)ne synthesis (Figure 6.1).

The ability to increase the yield of naturally produced alkanes in *Synechocystis* (Wang et al., 2013, Hu et al., 2013) along with their export out of the cell would be a step towards producing low-cost renewable transportation fuels. To date, the mechanisms for the extraction of synthesised alkanes in cyanobacteria are still not well established. However, there is a patent on the expression of an internal membrane protein that facilitates the export of produced alkanes in *Synechocystis* into the culture medium (Reppas et al., 2010). The next section

describes the AlkL transporter and its potential role in alkane transport in *Synechocystis*.

6.1.2 The AlkL transporter

The outer membrane of Gram negative bacteria acts as a protective layer that prevents both hydrophobic and hydrophilic substances from crossing freely into the cell. This can be attributed to the presence of the lipopolysaccharide layer on the outer membrane (Nikaido, 2003, Hearn et al., 2009). Several outer membrane proteins have been identified in various bacteria that are likely to aid in the transport of small hydrophobic molecules, such as the AlkL and OmpW family proteins (van Beilen et al., 2001, Hong et al., 2006). However, the only protein family currently confirmed to be involved in the uptake of hydrophobic substances is the FadL family in *E. coli* (Hearn et al., 2009). The *alkL* gene found in *Pseudomonas putida* GPo1 is part of the alkBFGHJKL and alkST operons carried on the OCT-plasmid (Chakrabarty et al., 1973). The two operons are reported to contain genes for enzymes responsible for alkane metabolism (van Beilen et al., 1992). The AlkL protein (230 aa) was found to show a 27% identity to the OmpW outer membrane protein of *E. coli* (Hong et al., 2006) which in turn is homologous to the OmpW protein of *Vibrio cholera* (van Beilen et al., 1992). A key distinct similarity between the OmpW and the AlkL membrane proteins, both of which are eight stranded beta barrels, is that they contain a hydrophobic channel rather than the hydrophilic interior typically found in membrane protein transporters (Hong et al., 2006). Another characteristic similarity is the lateral opening of the channel, which is hypothesised to facilitate the exit of hydrophobic compounds from within the channel into the membrane rather into the hydrophilic periplasm (Figure 6.2).

It has been previously hypothesised (van Beilen et al., 2001, Hong et al., 2006, Hearn et al., 2009, van Beilen et al., 1992) that given the location of the *alkL* gene in the alkane degradation operon and its position in the outer membrane, that it is involved in transporting alkanes into the cell and their subsequent degradation. Nevertheless, until recently, the precise function has remained unknown with no observed effect on the microorganisms growth rates when the *alkL* gene is deleted (van Beilen et al., 1992). The suggested explanation for the absence of an observed effect upon deletion of *alkL* is that the AlkL protein may only be functional when the microorganism is growing in conditions where the

levels of alkanes are very low and where the AlkL protein would be necessary for importing of the alkanes. It has also been proposed that the redundancy of the outer membrane transporters exhibiting the same function may be another possible explanation (Hong et al., 2006).

Work done by Grant (Grant, 2012) provided clear evidence on the role of the AlkL protein in transporting long-chain alkanes across the membrane. The mechanisms proposed by which the AlkL protein functions in both importing and exporting $>C_{12}$ alkanes are demonstrated in Figures 6.3 and 6.4, respectively. The mechanism by which the bio-pump is proposed to import hydrophobic compounds is known as lateral diffusion (Figure 6.3) and involves the facilitated binding of the hydrophobic molecules to the extracellular region of the AlkL protein. The binding happens with high affinity allowing the molecules to cross the lipopolysaccharide layer via hydrophobic channels to get into the cell (Hearn et al., 2009, Hong et al., 2006). Figure 6.4 illustrates the proposed method for product export where the products within the cell diffuse into the cytoplasmic membrane, cross the periplasm and into the outer membrane. Here diffusion is blocked due to the presence of the polar lipopolysaccharide layer. The AlkL protein forms a hydrophobic channel that allows the product to bypass the lipopolysaccharide layer and exit the cell. However, the hydrophobic nature of the outer membrane of the AlkL protein brings these products back into the cell until a dynamic equilibrium is reached. According to Hearn and colleagues. (2009), all transport proteins in which lateral diffusion is proposed, function as efflux pumps. Since lateral diffusion is proposed for the AlkL protein, we can therefore consider it a potential efflux pump that may allow for the export of alkanes from within the cell.

As mentioned above the expression of the bacterial origin AlkL in *Synechocystis* seems promising given its classification as a Gram-negative bacterium. However, it should be noted that the cyanobacterial cell wall shows many differences to other Gram-negative bacteria. As an example, the peptidoglycan layer found in cyanobacteria is considerably thicker than that found in most Gram-negative bacteria and the degree of cross-linking between the peptidoglycan chains also varies compared to those of other bacteria (Hoiczyk and Hansel, 2000). Details of the cyanobacterial cell structure can be found in chapter 1. These facts need to be kept in mind when conducting comparative studies between *E. coli* and *Synechocystis*.

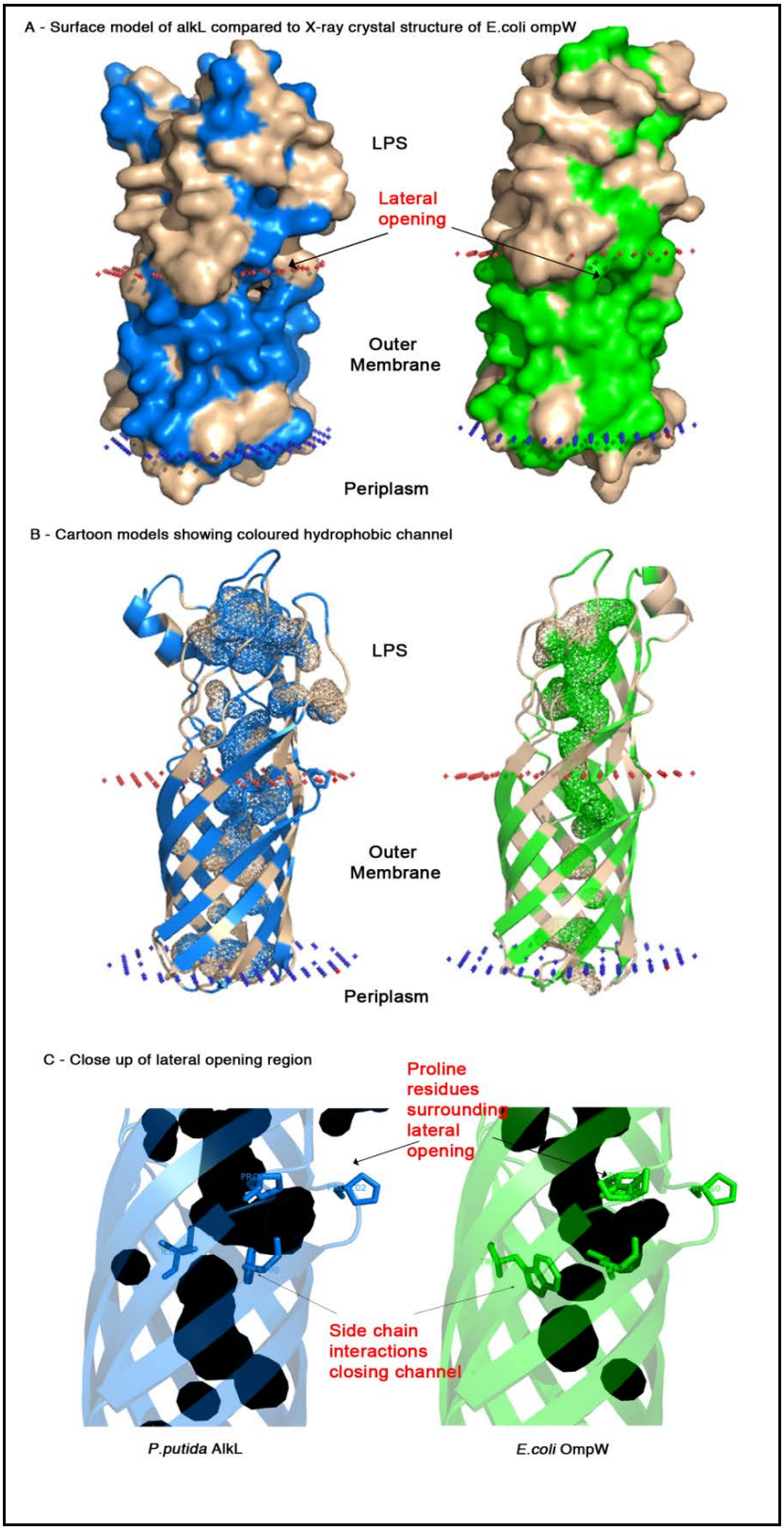


Figure 6.2: Structural comparison between the homology model of AlkL and the X-ray crystal structure for *E.coli* OmpW.

(A) Surface model indicating hydrophobic side chains for AlkL (blue) and OmpW (green). (B) Cartoon structure of AlkL model with cavities shown as a mesh. (C) Close up of the proline residues forming the lateral opening into the outer membrane and the side chain interactions closing the channel beneath the lateral opening. LPS; lipopolysaccharide layer (unpublished figure provided by C. Grant).

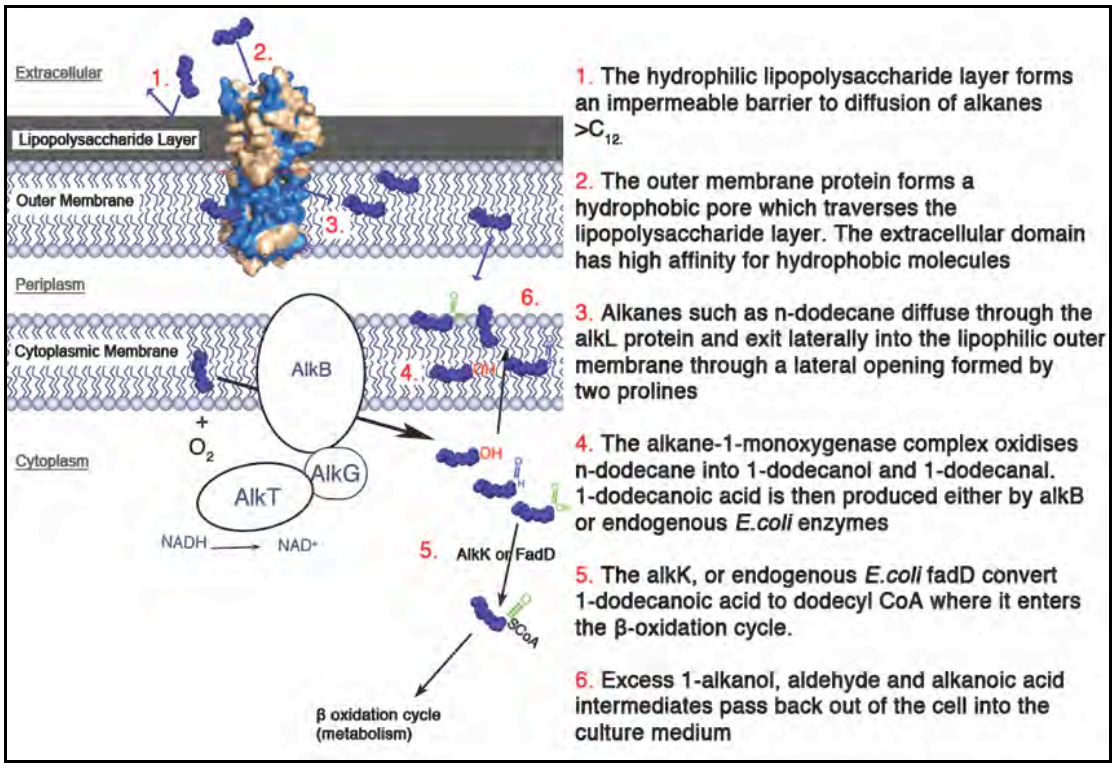


Figure 6.3: A schemetic representation of the proposed mechanism for alkane import.

The figure demonstrates the proposed mechanism by which the AlkL protein aids in the uptake of long chain alkanes across the lipopolysaccharide layer (LPS) of gram-negative bacteria (unpublished figure provided by C. Grant).

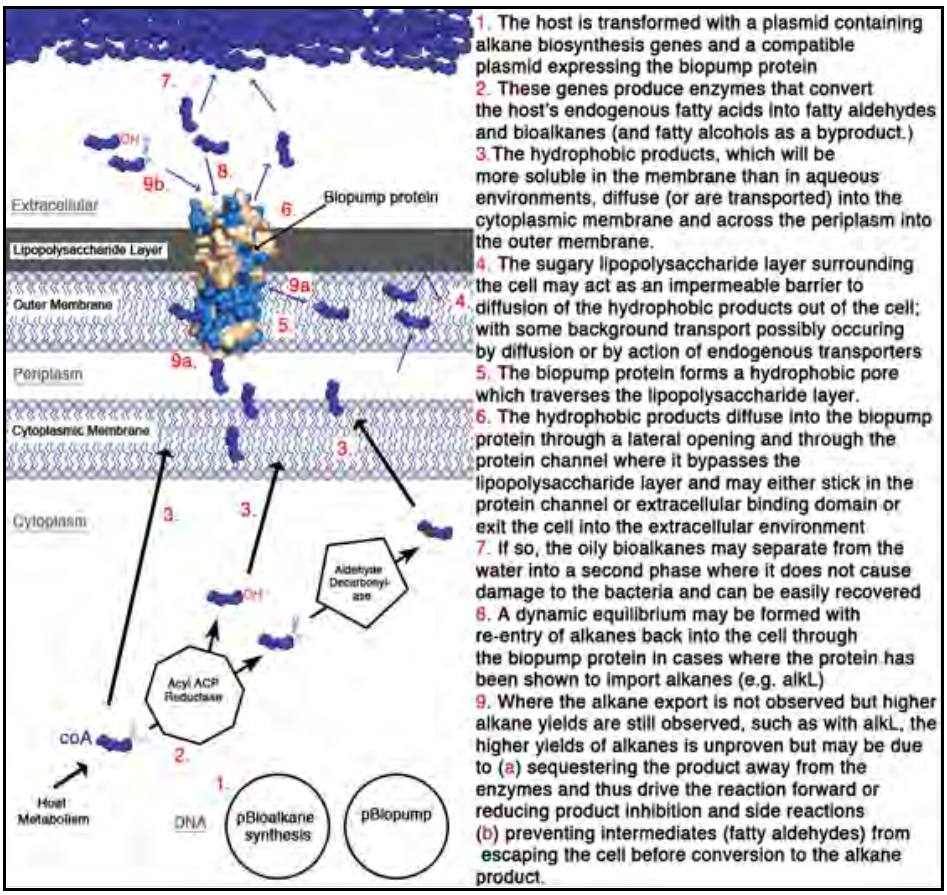


Figure 6.4: A schemetic representation of the proposed mechanism for alkane export.

The figure demonstrates the proposed mechanism by which the AlkL protein aids in the export of long chain alkanes across the lipopolysaccharide layer (LPS) of gram-negative bacteria (unpublished figure provided by C. Grant).

6.2 Aims

- The development of a bio-pump expression construct that would allow for the expression of the *alkL* gene in *Synechocystis* under the light regulated *psbA2* promoter and the nickel inducible *nrsB* promoter.
- Investigate the role of the expressed bio-pump in transporting hydrophobic compounds across the membrane of *Synechocystis* and test for increased alkane synthesis.

6.3 Results and discussion

6.3.1 Expression of *alkL* gene in *Synechocystis* under different promoters

6.3.1.1 Creating the pLAH.A2.alkL and the pLAH.nrsB.alkL plasmids

PCR was utilised for the amplification of the *alkL* gene from plasmid pJ811 (provided by Dr Grant) using primers that would add an HA tag sequence onto the 3' end of the *alkL* coding sequence, and *Nde*I and *BamH*I restriction sites (on the 5' and 3' ends respectively) for downstream cloning purposes. PCR conditions used were as described in section 2.3.5. Primers used to amplify the *alkL* gene and the full sequence of the pJ811 plasmid including the *alkL* gene can be found in Appendix 3 and Appendix 14 respectively. PCR results shown in Figure 6.5 confirm the amplification of the *alkL* gene (751bp).

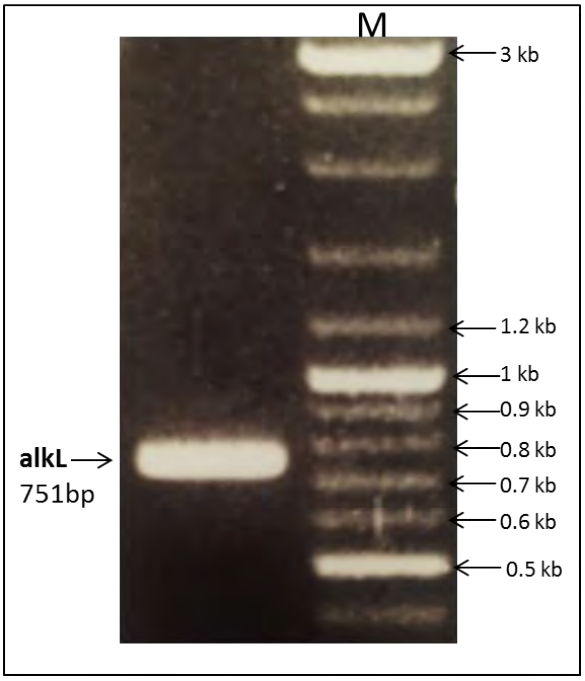


Figure 6.5: The amplification of *alkL* gene with an HA tag.

PCR amplification of the *alkL* gene from pJ811:71557–ALC plasmid template.

The purified PCR product was digested with *Nde*I and *Bam*HI before being ligated into pLAH.A2 and pLAH.nrsB expression vectors digested with the same enzymes (section 5.3.1 chapter 5). Details for enzyme digestions and ligation can be found in sections 2.3.11.2 and 2.3.11.3 respectively. The ligation mix was then transformed into competent *E. coli* cells (section 2.4.1) and selected on LB plates supplemented with kanamycin added to a final concentration of 50 µg/ml. No transformants were obtained when the *alkL* gene was ligated to the pLAH.A2 expression vector while many transformants were recovered from the ligation of

the *alkL* gene in pLAH.nrsB. One possible explanation for the absence of *E. coli* transformants for the pLAH.A2.alkL construct, despite several attempts, might be that expression of the *alkL* gene in the *E. coli* is toxic and leads to cell death. Membrane proteins are typically known to be toxic to the cell when over expressed (Narayanan et al., 2011, Akiyama, 2009) and the AlkL protein specifically has been shown to have a highly detrimental effect on cell viability when expressed in *E. coli* (Grant, 2012). Furthermore, we would expect the *alkL* to be expressed in *E. coli* under the *psbA2* promoter (strong light regulated promoter) as this promoter has previously shown to successfully express genes in *E. coli* (as seen with the expression of the toxic *GES* gene in *E. coli* in chapter 5).

Test digests performed on isolated plasmid obtained following the ligation of the *alkL* gene into pLAH.nrsB showed the successful insertion of the *alkL* gene. In order to confirm that the HA tag sequence was successfully attached to *alkL*, aliquots of the purified plasmids were sent for sequencing (section 2.3.6) along with primers used to check for the correct insertion of *alkL* in the *psbA2* locus (Appendix 3). DNA sequence results confirmed the presence of the HA tag sequence at the end of the *alkL* coding region (Appendix 15).

6.3.1.2 *Synechocystis* transformation with pLAH.nrsB.alkL

The pLAH.nrsB.alkL plasmid obtained in section 6.3.1.1 was used to transform WT *Synechocystis* as described in section 2.4.2. Colonies appeared within one week of selection on medium supplemented with 200 µg/ml kanamycin (Figure 6.6.A). Genomic DNA was isolated from four colonies after they had been re-streaked three times on fresh selective medium. The isolated DNA was analysed by PCR using primers (located upstream of the *psbA2* gene and other within the pLAH.nrsB vector) designed to check for correct integration of *alkL* gene in the *psbA2* locus (Appendix 3). PCR results shown in Figure 6.6.B confirm the successful integration of the *alkL* gene in the *psbA2* locus under the control of the *nrsB* promoter.

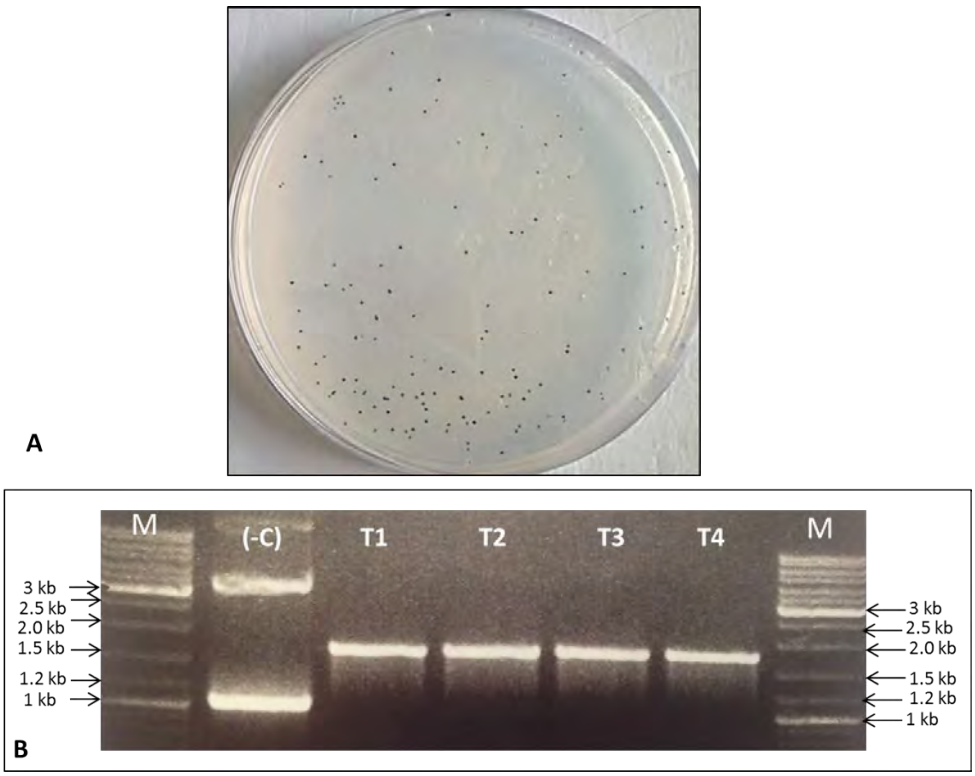


Figure 6.6: Transformation of *Synechocystis* with pLAH.nrsB.alkL construct and PCR confirmation.

(A) Many transformants were obtained from transformation of WT *Synechocystis* with pLAH.nrsB construct expressing the *alkL* gene. (B) The successful isolation of *Synechocystis* transformants containing the *alkL* gene (~1.7 kb product). Negative control: pLAH.nrsB plasmid template (~1 kb product). A non-specific band appears at ~3.1 kb.

6.3.1.3 Western blot analysis to test for the synthesis of the HA-tagged AlkL

In order to test for the expression of AlkL protein under the *nrsB* promoter in *Synechocystis*, western blot analysis was carried out using anti-HA antibodies. The western blot was set up as described in chapter 2 (section 2.6). The synthesis of the AlkL protein (~25 kDa) was observed in cultures that were pre-induced for four hours with $NiCl_2.6H_2O$ added to a final concentration of 6.4 µM (Figure 6.7.A, lane 2). Un-induced cultures showed no leaky expression of *alkL* confirming the successful and tightly controlled expression of the gene under the *nrsB* promoter in *Synechocystis* (Figure 6.7.A, lane 1). A positive control represented by an engineered strain of *E. coli* expressing the AlkL protein with a C-terminal 6xHis tag was used to confirm protein detection and can be seen in Figure 6.7 (B). The positive control was provided by C. Grant.

Interestingly, the AlkL protein expressed in *Synechocystis* appears larger (~25 kDa) than that expressed in *E. coli* as reported by both Grant et al., (unpublished data) and (Kok et al., 1989, Eggink et al., 1987) who obtained protein sizes of ~22 kDa and 20

kDa, respectively. Nevertheless, our finding is in agreement with van Beilen. (1992), who also reported an AlkL protein size of 25 kDa. It can be suggested that since the AlkL protein has three closely spaced potential peptidase cleavage sites (van Beilen et al., 1992) and if the protein was cleaved at the site with the highest score according to (von Heijne, 1986), the calculated molecular mass of the mature AlkL would be 22 kDa. Once the additional size of the HA tag (~1.1 kDa) is included, then the mature size is reasonably close to the observed molecular mass of 25 kDa.

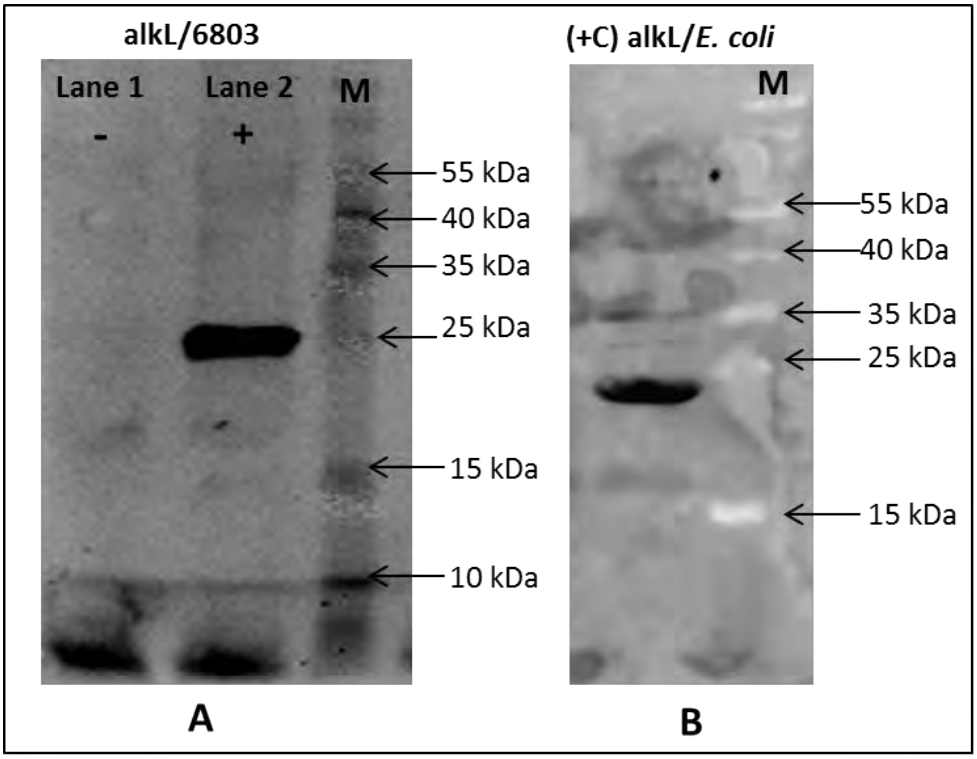


Figure 6.7: The induced expression of *alkL* gene product in *Synechocystis.*

(A) The *alkL* gene product was successfully expressed in *Synechocystis* (~25 kDa) after the induction of the *nrsB* promoter. (B) *E. coli* was used as a positive control to demonstrate the expression of *alkL* gene (~22 kDa). Anti-HA antibodies were used in (A) while anti-His used in (B).

The ability to regulate the expression of the *alkL* gene was previously identified to be critical for stable accumulation of AlkL in *E. coli*, probably as a consequence of its toxicity on the cell when over expressed. It was therefore necessary to have a way to regulate the expression of *alkL* in *Synechocystis*. This was done by expressing the *alkL* gene under the tightly controlled *nrsB* promoter and using different inducer concentrations to achieve titratable expression levels. For this, another western blot was set up, as detailed in section 2.6, where the *alkL* expressing strain was grown and induced for four hours at low and high concentration of the inducer (0.64 µM and 6.4 µM $NiCl_2.6H_2O$, respectively) and

the detection was done using the Odyssey system that allowed for band intensity measurements to be made for relative comparisons. It is clear from the western blot results presented in Figure 6.8 that the level of protein correlated with the inducer concentration with higher protein expression levels achieved through higher induction while no protein was expressed in the un-induced culture.

In *Synechocystis*, the level of *alkL* gene product was approximately 2.5 times higher when induced with the higher concentration of nickel (6.4μM) (Figure 6.8 lane 3). To check for the reproducibility of the results, an experiment was set up where three biological samples were tested for controlled expression of the *alkL* gene. Results presented in Figure 6.9 show that induced cultures of the three biological samples (represented by a + sign) showed a protein product at the expected size (~25 kDa). However, faint bands also appear in the same location in the un-induced cultures. This could either indicate a leaky expression of the *alkL* gene or may just be due to the presence of non-specific bands (several faint non-specific bands appear in the blot). In order to quantify the amount of the AlkL protein expressed in the cells, 50 ng and 100 ng of the standard HA-tagged Human CARHSP1 protein were loaded with the samples in Figure 6.9. The quantification was not successful since the standard protein was degraded probably due to thaw-freeze effects and therefore the quantification needs to be repeated.

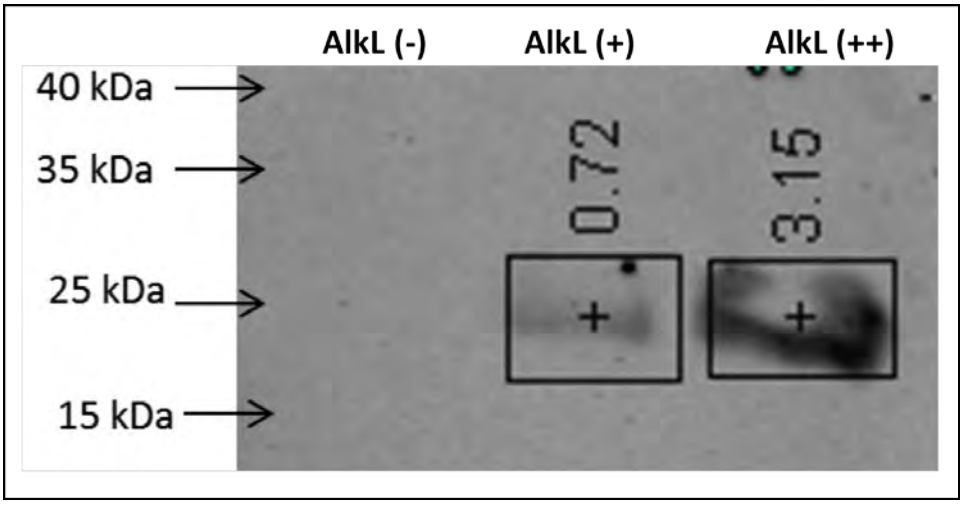


Figure 6.8: The titrated expression of the alkL gene in *Synechocystis.*

Controlled expression of the *alkL* gene was achieved by inducing *Synechocystis* cultures with low (0.64 μM) represented by (+) and high (6.4 μM) (represented by ++) concentrations of $NiCl_2.6H_2O$. An un-induced culture was used as a negative control (-).

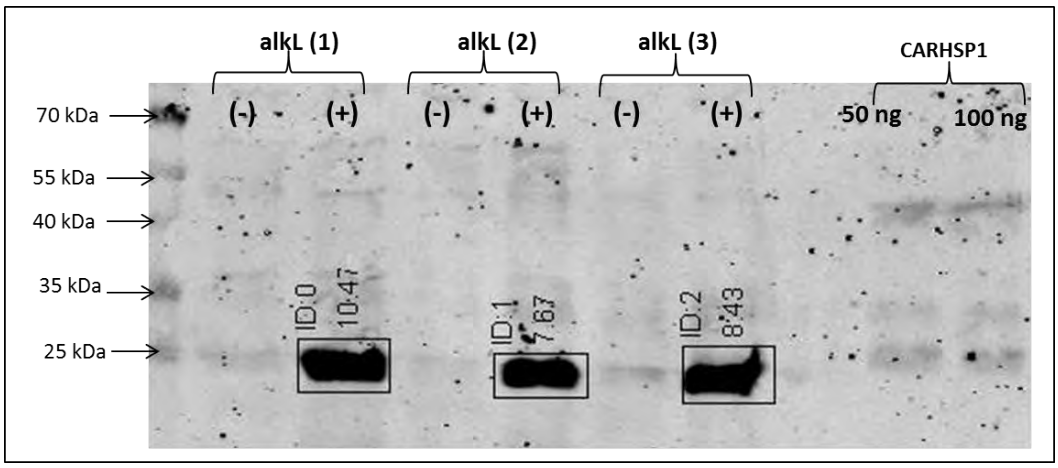


Figure 6.9: Quantification of the *alkL* gene product.

An attempt to quantify the *alkL* gene product using the Odyssey system and CARHSP1 protein. The CARHSP1 protein (50 ng and 100 ng) was degraded and quantification was not achievable. (-) un-induced cultures, (+) cultures induced with 6.4 µM of $NiCl_2.6H_2O$, alkL (1-3) show three independent strains expressing the alkL protein.

6.3.2 Phenotype analysis of the AlkL expressing strain

Based on the proposed mechanism of alkane transport of the AlkL protein suggested in Figures 6.3 and 6.4 (lateral diffusion and equilibrium based), the effect of expressing the AlkL on intracellular heptadecane production and recovery in *Synechocystis* was investigated. Furthermore, the effect of using an auxiliary phase (n-dodecane) along with a detergent (Triton X100) was examined. N-dodecane was added to the culture as an exogenous alkane to test whether; (i) AlkL is functional in translocating it inside the cell, (ii) It will improve the yield and recovery of endogenous alkanes by milking it outside the cell. On the other hand, Triton X100 was used to aid recovery of alkanes that may adhere to the cell membrane. Moreover, in *E. coli*, the combination of the detergent and auxiliary phase has shown to improve recovery of C_{17} alkanes (Grant et al. unpublished).

To test these parameters on *Synechocystis*, an experiment was set up as illustrated in Table 6.1. In this experiment, both the AlkL expressing strain and the negative control (*Synechocystis* transformed with the pLAH.nrsB expression vector) were grown to an OD_{750}=0.6 before being added into the wells of the polypropylene plates in the presence or absence of the inducer ($NiCl_2.6H_2O$), n-dodecane and Triton X-100. Cultures were allowed to grow for 21 hours (to reach mid log phase) before being analysed phenotypically and later harvested for GC analysis (section 6.3.3).

Table 6.1: The set up used for phenotype analysis of AlkL expressing strain

	Triton 0.01%	1	2	3	4	5	6	
A	+	3 ml	3 ml	3 ml	3 ml	3 ml	3 ml	0% n-dodecane
B	-	3 ml	3 ml	3 ml	3 ml	3 ml	3 ml	
C	+	3 ml	3 ml	3 ml	3 ml	3 ml	3 ml	9% n-dodecane
D	-	3 ml	3 ml	3 ml	3 ml	3 ml	3 ml	
		0µM	0.64µM	6.4µM	0µM	0.64µM	6.4µM	inducer concentration
			alkL			negative control		strain

6.3.2.1 Phenotype analysis:

The cultures grown according to the set up presented in Table 6.1 displayed a distinct phenotype that can be clearly seen in Figure 6.10.B. The effect of expressing the *alkL* gene at high concentrations of the inducer (6.4 µM) can be seen in column three where the cultures displayed a pale green color indicating reduced pigmentation. C3, which contained both n-dodecane and Triton X100, was an exception as it maintained a green colour with very little loss of pigmentation when compared to the negative control (C1). The loss of pigmentation is unlikely due to the inducer as the negative controls of the equivalent set up, shown in column 6, showed no detrimental effect on the cells. This observation clearly indicates that the lower level of pigmentation was a result of the expression of the *alkL* gene in *Synechocystis*. However in order to confirm that the phenotype change in the *alkL* expressing strain was due to the function of the AlkL protein in accumulating intracellular alkanes and not due to weakening of the cell membrane as a result of over expression of the protein, measurements of intracellular levels of alkanes were done in section 6.3.3.

The effect of adding n-dodecane in the presence of Triton X-100 can be seen in row C (C1 to C6) where cultures exhibited a darker green culture indicating healthier cells. It is speculated that this may be a result of (i) an increased gas exchange in the culture since gases are known to be 10 fold more soluble in n-dodecane compared to water (Ngo and Schumpe, 2012) (ii) n-dodecane might be extracting a toxic metabolic byproduct away from the cell resulting in healthier cells and increase in the chlorophyll content or (iii) n-dodecane is being used as a

carbon source for growth. It is interesting to note that the healthier appearance of the cells was only observed when Triton was added to n-dodecane (compare rows C and D) suggesting that the detergent is necessary for the role of n-dodecane on the cell. It was observed that higher cell densities in *E. coli* were achived when cells where grown in the presence of n-dodecane irrespective to the presence of the AlkL protein (Grant et al., 2012). Cell density measurements of *Synechocystis* were carried out to compare the results to those obtained for *E. coli* and is discussed in the following section.

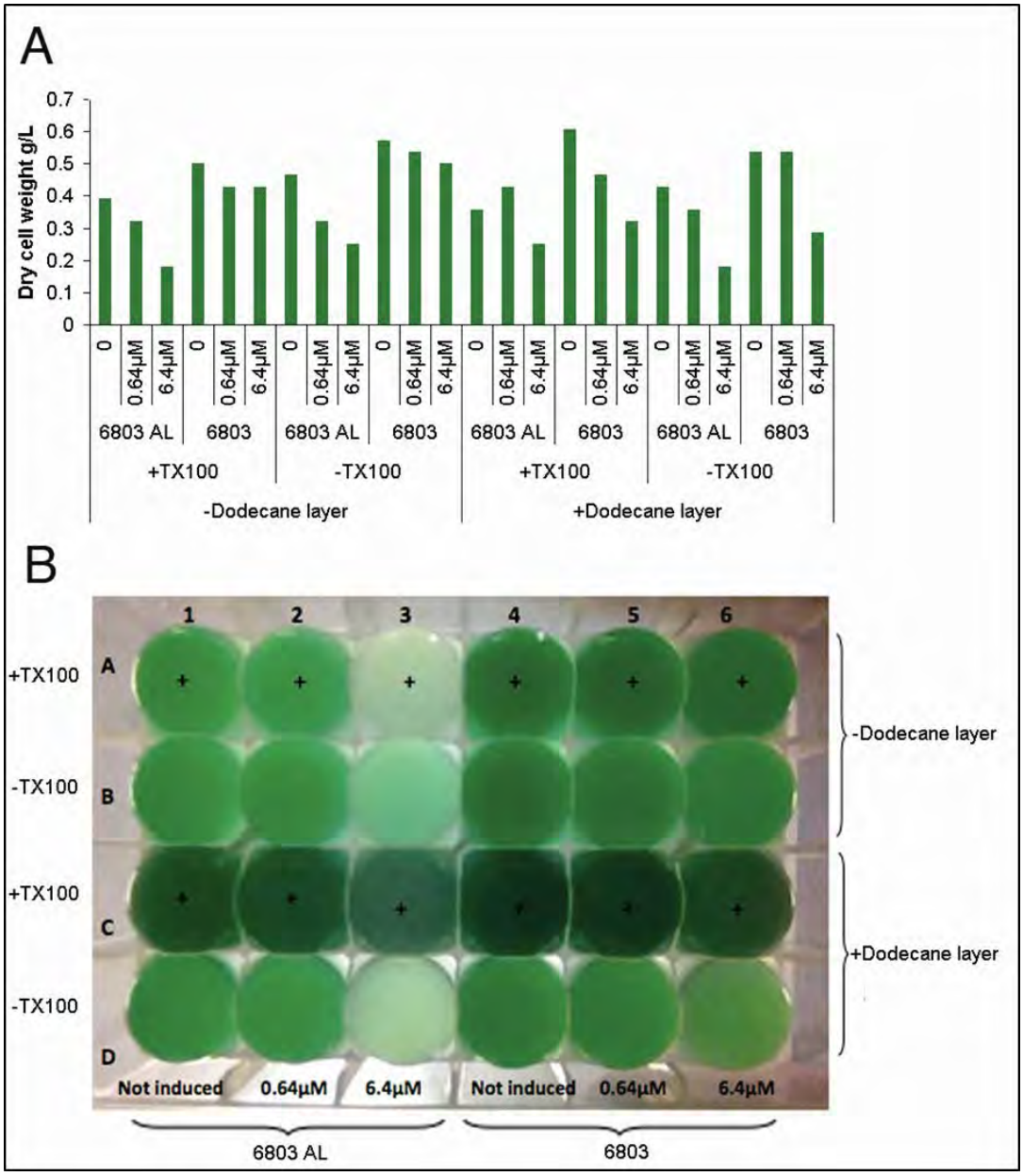


Figure 6.10: Phenotype analysis of the AlkL expressing strain.

(A) The cell dry cell weight is inversely proportional to the inducer concentration. The AlkL expressing strain generally had a lower dry cell weight as compared to its equivalent WT strain. (B) The AlkL expressing strain of *Synechocystis* showed a clear phenotype change when induced with 6.4 µM of $NiCl_2.6H_2O$ under all four conditions tested. + and – represent presence or absence of n-dodecane or tritonX100.

6.3.2.2 Cell density determination:

The cell density was measured as described in section 2.10 and results obtained are presented in Figure 6.10.A. The results show that expression of the AlkL protein in *Synechocystis* was inversely correlated with cell density, which provides some evidence that the AlkL protein was being expressed in a functional form and this behavior matches what has been demonstrated in *E. coli*. On the other hand, WT strains also showed a reduction in the dry cell weight with increased inducer concentration under all conditions tested. Nevertheless the reduction was on a much smaller scale and was most likely a result of prolonged induction. Interestingly, under all conditions tested in Figure 6.10.A, the dry cell weight of the un-induced *alkL* expressing strain was found to be less than that of the corresponding WT strain indicating that the presence of the *alkL* gene adds a metabolic burden on the cell.

Next we wanted to focus on the effect of the AlkL protein on dry cell weight measurements independent of the presence of a second-phase, for this, rows A and B from Figure 6.10.B (absence of n-dodecane) were examined carefully and re-presented in Figure 6.11. Figure 6.11.A shows that only the AlkL expressing strains of *Synechocystis* (6803 AL) induced with 6.4 µM of $NiCl_2.6H_2O$ showed a clear phenotype change represented by the pale green culture as opposed to dark green color of the rest. This was the case in both cultures grown in the presence or absence of triton indicating that the presence of triton alone has no apparent effect on the cells. It is important to note that although the phenotype appearance of the cultures corresponds fairly well with the dry cell weight measurements in a single phase system (a paler color indicated reduced dry cell weight) (Figure 6.11). The culture colour cannot be used to predict the effect on the dry cell weight when using a two-phase culture system. This is evident in Figure 6.10.B where the dark green colour of cultures grown in the presence of n-dodecane (row C) did not translate to higher dry cell weight measurements (6.10.A). This result can be used to rule out the possibility of n-dodecane used as a carbon source for growth. The following section describes GC analysis conducted to examine the intracellular levels of alkanes in AlkL expressing strains.

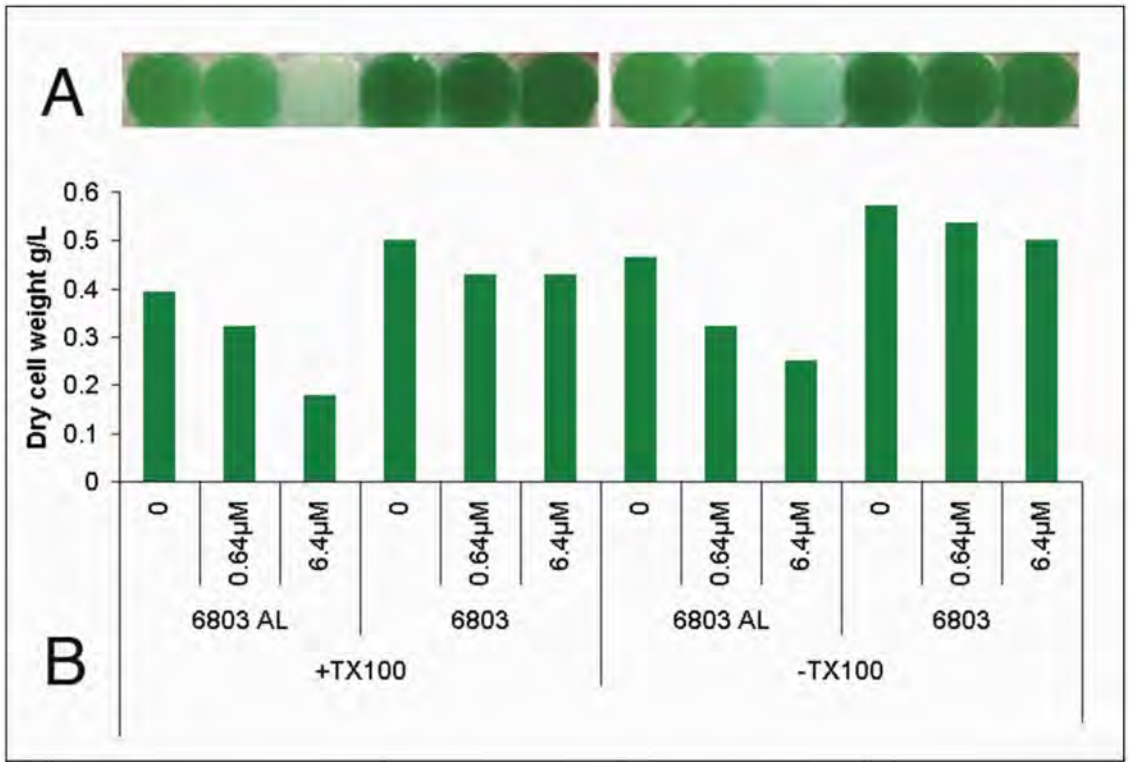


Figure 6.11: Effect on the dry cell weight of expressing *alkL* in a single phase.

(A) Only AlkL expressing strains of *Synechocystis* (6803 AL) induced with 6.4 µM of $NiCl_2.6H_2O$ showed a clear phenotype change in both cultures grown in the presence or absence of Triton. (B) AlkL expressing strains had less dry cell weight than the WT strains while the inducer concentration showed an inversely proportional relationship with the dry cell weight of the AlkL expressing strain.

6.3.3 GC analysis of *alkL* expressing strains

For the GC analysis, three biological replicates of the AlkL expressing strains were cultured in the same set up shown in Table 6.1 and displayed similar phenotypes to that seen in Figure 6.10. The pellet and supernatant extracts of the cultures (section 2.8.2) were analysed by GC (section 2.8.4) for the detection and quantification of the naturally produced C_{17} heptadecane and the exogenously added C_{12} alkane, n-dodecane.

This was done to compare (i) the proportion of heptadecane in the pellet and supernatant fractions under different conditions (in particular the effect of *alkL* expression), (ii) compare the uptake of n-dodecane into the pellet for rows C and D.

Overall production of heptadecane: GC measurements of the average amount of heptadecane in *Synechocystis* expressing the AlkL protein presented in Figure 6.12 showed that the induced expression of the AlkL protein decreased the overall yield of heptadecane under all conditions tested (including the presence of a two-phase culture system). However, strains lacking the AlkL protein (negative controls) also showed the same result indicating that the reduced intracellular levels of the heptadecane was not due to the expression of the AlkL protein but more likely a side effect of prolonged nickel induction. This is contrary to what was observed in *E. coli* where the expression of the AlkL protein resulted in an increased yield of alkanes. Furthermore it was reported that the alkane yields were enhanced in *E. coli* by the addition of n-dodecane and detergent (Triton X100) even in the absence of the AlkL transporter. Grant. (2013), hypothesised that this might be due to the produced alkane being milked out of the cell, via naturally occurring transporters, into the n-dodecane layer reducing its concentration within the cell and driving the reaction forward. This is not probable in *Synechocystis* since, to the best of my knowledge, there is no natural exporter of alkanes.

Export of alkanes: Looking at the average C_{17} content of supernatant and pellet fractions of three biological replicates shown in Figure 6.12, we can see that there is no observed increase in the export of the produced heptadecane in induced AlkL strains grown in single phase. We can therefore conclude that the AlkL transporter did not enhance product recovery in this case.

Auxiliary phase system and C_{17} export: GC results for influx of n-dodecane (C_{12}) into the cell were inconclusive due to experimental errors that resulted in a high signal to noise ratio and were therefore not presented in this work. However, GC measurements of the proportions of heptadecane in the supernatant and pellet presented in Figure 6.12 shows that the addition of the n-dodecane auxiliary phase resulted in the increase of heptadecane secretion. This is similar to the results observed in *E. coli* cells where the presence of n-dodecane second-phase resulted in the extraction of intracellular alkanes irrespective of the presence or absence of the transporter. However, contrary to what was observed in *E. coli*, the detergent did not seem to enhance alkane recovery in the second-phase.

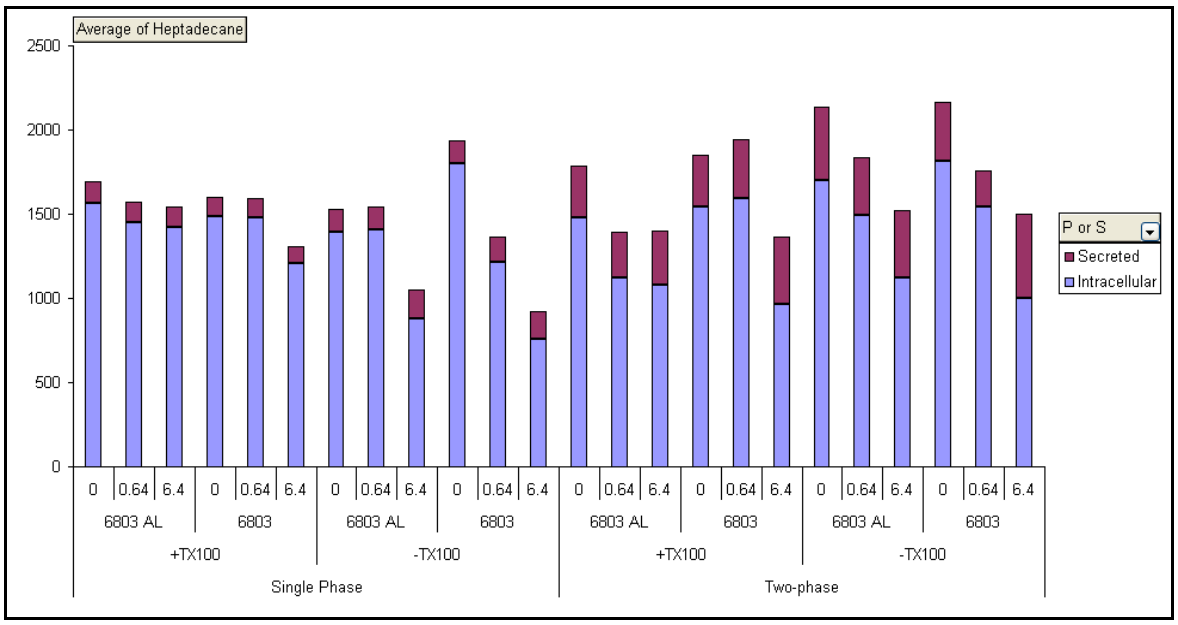


Figure 6.12: GC-FID analysis of Heptadecane in *alkL* expressing strains.

Heptadecane production was decreased in both WT and AlkL expressing strains upon induction under all conditions tested. The figure represents the average of three replicates.

6.4 Conclusion and future work

The precise function of the AlkL outer membrane protein remains unknown despite a general agreement that the Alk operon (including the *alkL* gene) is responsible for alkane uptake and metabolism (van Beilen et al., 2001, Hong et al., 2006, Hearn et al., 2009, van Beilen et al., 1992). The function was proposed according to the general structure of the AlkL transporter (distinct hydrophobic core) and its location in the outer membrane, and both features support the transport of hydrophobic molecules across the outer cell membrane. Analysis of an *E. coli* strain engineered for alkane production demonstrated an increase in both the uptake and accumulation of alkanes (Grant, 2012). Work presented in the first part of this chapter confirmed the controlled and tightly regulated expression of the *alkL* gene, for the first time, in *Synechocystis* under the control of the *nrsB* inducible promoter system (Figures 6.7 and 6.8). However, attempts to express the *alkL* gene under the strong light regulated promoter (*psbA2)* failed, probably due to the expression of the gene in *E. coli* cells that resulted in cell death and made subsequent cloning steps unachievable.

Cyanobacterial strains expressing the *alkL* gene were analysed phenotypically under various conditions and results, presented in Figure 6.10.B, clearly showed that the expression of the *alkL* gene led to an observed phenotypic difference represented by reduced pigmentation indicative of the probable functionality of

the AlkL protein. Furthermore, it was evident that the presence of n-dodecane resulted in healthier cells represented by the darker green culture, probably as a result of an increased gas mass transfer in the presence of n-dodecane and an increase in the chlorophyll content of the cells (not tested).

The next steps included examining the effect of expressing the AlkL protein on (i) the dry cell weight of the strain (ii) the accumulation of total alkane (iii) the export of the intracellular heptadecane (iiii) the effect of growing the culture in an auxiliary phase in the presence or absence of a detergent.

Results from GC analysis suggested that the dry cell weight of AlkL expressing strains was inversely correlated with the inducer concentration suggesting that the AlkL protein was probably functional (Figure 6.11). On the other hand, GC analysis conducted on three biological samples of AlkL strains showed a reduction of the total yield of heptadecane (Figure 6.12) however since the negative controls (strains lacking the *alkL* gene) also showed reduced yield of heptadecane, probably due to the effect of nickel induction, the role of the AlkL transporter in *Synechocystis* is still not identified. Furthermore, the same figure showed that the expression of the AlkL protein did not result in the export of the heptadecane as the proportion of C_{17} in the supernatant was not improved post *alkL* induction. Unfortunately the GC results of the C_{12} n-dodecane analysis were not conclusive and should be repeated in order to confirm this hypothesis. There was variability in the total alkane yield between the three biological samples presented in Figure 6.12 and trying to find a possible explanation may be challenging. Hu et al.(2013), also reported variable amounts of accumulated heptadecane (C_{17}) between different samples of wild type *Synechocystis*. One possible explanation is that microbial hydrocarbon synthesis largely depends on the growth conditions (Ladygina et al., 2006). Despite growing the cultures under the same conditions, small variations within the microplates and slight differences in the light intensities might have led to such variations. Recommended future work would include the expression of small amounts of the AlkL protein and its quantification in order to achieve the desirable induction level that would allow for the expression of the protein without resulting in cell toxicity. This will help identify the role of the protein avoiding concerns related to possible cell damage due to over-expression of the protein. Furthermore the GC set up needs to be repeated to examine the transport of n-dodecane across the cell in order to identify its role in enhancing growth conditions of *Synechocystis* and possible export of alkanes.

Chapter 7: GENERAL DISCUSSION

7.1 Main Findings

7.1.1 Attempts for the development of a simple system for random insertion mutagenesis in *Synechocystis* 6803

The creation of random mutants in *Synechocystis* is currently based on classical mutagenesis methods such as chemical and UV mutagenesis, both of which result in difficulties in the identification of the mutation in the genome (Labarre et al., 1989). The use of transposons for mutagenesis is another classical mutagenesis method but is of limited use in cyanobacteria due to its low efficiency (Golden, 1988). To overcome the limitations of these classical forward genetics methods, an insertional mutagenesis method was developed by (Labarre et al., 1989) where the disruption of the host genome is done by the insertion of an antibiotic cassette isolated from a bacterial plasmid that is randomly ligated to a pool of restriction fragments of the host genome, and used directly for transformation into *Synechocystis.* Nevertheless, this method still relies on multiple digestion and ligation steps prior to transformation, making it a long process. Therefore, in chapter 3, I conducted experiments that aimed to develop a simple and rapid method for random foreign gene insertion into the genome of *Synechocystis* that would not require the construction of cloning vectors, digestions or ligation steps and which does not rely on homologous recombination to achieve transformation. The idea was to ultimately use the proposed method for identification of novel mutants using forward-genetic screens. To achieve this goal, it was decided to utilised a *recJ*– strain of *Synechocystis* as a recipient strain for the transformation of a 'naked' km^R cassette (not flanked by any DNA homologous to the cyanobacterial genome). The reason being that it appears that the loss of the RecJ exonuclease in the *recJ*– strain extends the half-life of the transforming DNA allowing inefficient illegitimate (*i.e.* non-homologous) recombination events to occur.

Results presented in chapter 3 confirm the successful transformation using naked km^R cassettes (isolated as both blunt and sticky end cassettes) in the Δ*recJ*.ble *Synechocystis* mutant. On the other hand, Southern blot analysis conducted on 16 isolated transformant lines revealed that the integration of the km^R cassettes were at a specific location in the host's genome and not random, as predicted. Detailed investigation as to where the km^R cassettes were being integrated was done, and it demonstrated the integration of the cassettes at the *speA*/*psbA2*

locus. Furthermore, results confirmed that the integration was accompanied by a gene deletion where the km^R cassette was replacing the majority of the *psbA2* locus. The observation that genomic DNA had been deleted upon transformation with a km^R cassette in *Synechocystis* is similar to earlier findings by two groups (Chauvat et al., 1989, Labarre et al., 1989) who showed that transformants obtained using km^R and cm^R cassettes that had been ligated randomly to fragmented *Synechocystis* DNA, were usually accompanied by a gene deletion.

This led to two hypotheses to explain the unexpected result; the first was that the preferential integration of the km^R cassettes at the *speA*/*psbA2* locus was a result of something related to the sequence of the cassette. This hypothesis is supported by two facts: (1) the promoter and termination sequences of kanamycin are recognised by *Synechocystis* (Vermaas, 2007). Therefore, it is possible that this may have influenced its integration especially since the integration of the cassette happened between the promoter and terminator sequences of the deleted *psbA2* gene suggesting possible recognition of these sequences. Secondly, parallel work conducted in chapter 4 showed that the km^R cassette flanked on either end with short regions of homology to the genome (30 bp) also resulted, on some occasions, in integration at the *speA*/*psbA2* locus suggesting that when the short flanks linked to the km^R cassette are not identified by the host, the integration of the cassette is effected by the sequence itself. Conversely, blast results of the km^R cassette against the genome of *Synechocystis* showed no evidence of sequence homology between the cassette and the recipient strain.

The second hypothesis suggests that the integration of the km^R cassette at the *speA*/*psbA2* locus was due to specific sequences at this location making them act as “hot spots” for integration. This idea is supported by (Zang et al., 2007) who explained that the presence of “hot spots” in the genome of *Synechocystis* was attributed to the presence of a palindromic element in the genome at these sites which results in site specific recombination. These elements are identified as HIPI palindromic sequences and IS-like elements (Ikeuchi and Tabata, 2001). Examining the *speA*/*psbA2* locus for HIPI palindromic sequences revealed the presence of two of these sequences both upstream and downstream of locus suggesting that these sequences might have indeed resulted in the recombination event that took place in the transformants.

Experiments were performed to further examine the first hypothesis (integration influenced by the sequence of the km^R cassette) where the km^R cassette was replaced with a *ble* cassette and used in transformation of a Δ*recJ*.km mutant. Again, many successful transformants were obtained and Southern blot analysis again confirmed targeted integration of the cassette, emphasizing the inability to generate random integration using the proposed method. More importantly, targeted integration was at two distinct locations of the genome and since we previously predicted the *speA*/*psbA2* locus to be a possible "hot spot", we expected one of the two integration sites of the *ble* marker to be the *speA*/*psbA2* locus (as with the km^R cassette). Interestingly however, the *ble* marker was in fact being inserted in two different unidentified locations of the genome. These findings are very intriguing and suggest that the site-specific integration of the selectable markers is most likely affected by their sequence. However, in order to make a solid conclusion on the reason behind the site-specific integration of the selectable markers, further analysis needs to be done in order to identify the two locations at which the *ble* marker was being integrated and examine the presence of palindromic elements at the site of integration. It is possible that both the sequence of the selectable marker used and the presence of palindromic elements played a role in these site specific integration patterns.

It important to note that the experimental findings consistently showed that the *recJ*- strains (both the Δ*recJ*.ble and Δ*recJ*.km^R mutants) were more efficient than wild type strains in generating transformants, where wild type strains constantly failed to generate true transformants with any of the two selectable markers tested. This result is in agreement with findings reported by Kufryk et al. (2002), who demonstrated a two-fold increase in the transformation efficiency of *recJ*- strains as opposed to wild type strains.

In conclusion, the *recJ*- strains used in this study allowed for the integration of "naked" DNA into the host genome. However, results confirm that the site-specific integration of the selectable markers that took place is probably influenced by special genomic sequences (*e.g.* HIPI palindromic element) distributed throughout the genome or by the sequence of the selectable marker used for transformation or possibly by a combination of both. Therefore, it can be concluded that using *recJ*- strains to achieve the integration of heterologous DNA randomly across the *Synechocystis* genome does not look feasible.

Recommended future work would be, (i) the identification of the two integration sites of the *ble* selectable marker, (ii) examining the site of integration for the presence of genome repeats that may have led to targeted integration and (iii) test the integration pattern of a range of selectable markers. Results obtained from these experiments will help concluding whether the integration is based on the sequence repeats in the genome or rather related to the choice of the selectable marker used. If a selectable marker that does not lead to targeted integration can be identified the method might still be achievable.

7.1.2 Development of a 'one-step' PCR-based technique for insertion of foreign DNA into *Synechocystis* 6803

The most common method for targeted insertion of foreign DNA into *Synechocystis* is the construction of a transformation plasmid using traditional cloning methods such that a marker flanked on either side by a large (> ~500 bp) tract of homologous sequence (Golden et al., 1987, Vermaas, 2007). This method is time-consuming as it involves multiple cloning steps. Despite the recent development of next generation cloning methods that allow for the faster assembly of plasmids (Hillson et al., 2011), these new methods often require the creation of multiple PCR products and the use of reagents for their assembly into either an expression vector or the final PCR product used for transformation. Therefore, the aim of chapter 4 was to develop a rapid and convenient single-step method for creating gene knockouts or knock-ins in *Synechocystis* that avoids creating a transforming plasmid or the creation and ligation of multiple PCR products. This was to be achieved by firstly identifying the shortest homologous recombination flank size necessary for an efficient double homologous recombination transformation to occur in *Synechocystis* and subsequently use this information to generate gene knockouts in *Synechocystis* using a quick one-step PCR-based method for gene deletion.

The minimum flank necessary for transformation was identified after the successful isolation of positive *psbN* gene knockouts using a PCR product with 30 bp flanking sequences on either side of the km^R cassette. This result shows, for the first time, that a successful double homologous recombination event can take place in wild type *Synechocystis* using flanks less than 100 bp (albeit with

low efficiency) since the minimum flank size reported was 100-400 bp (Vermaas, 1998).

This information was used to design “long tail” primers that were successfully employed to create a *psbN*::km^R knockout PCR product in a single step, and the subsequent transformation resulted in the isolation of many kanamycin resistance colonies. Unfortunately, most transformants were not targeted *psbN* knockouts and further investigations showed that the PCR products were being inserted randomly into the genome via non-homologous recombination. After analyzing colonies obtained from transformations using PCR products from the traditional method and the one-step PCR method, it was apparent that the former generated from the traditional method resulted in targeted gene knockouts (homologous recombination, HR) most of the time, while PCR products for the same gene knockout from the one-step PCR method resulted in random or non-homologous recombination (NHR).

Therefore, to improve the HR:NHR ratio many parameters were tested. Results showed that using ssDNA for transformation and varying both DNA or cell concentrations did not lead to an improved HR:NHR ratio. However, replacing wild type *Synechocystis* with the Δ*recJ.*ble strain, as recipient strain for transformation of PCR products from the one-step PCR method, resulted in the isolation of 60 positive *psbN*::km^R knockouts indicating that the HR:NHR ratio can be significantly improved by simply using the more efficient Δ*recJ.*ble strain for transformation. However, these results were not reproducible suggesting that the strain used may have not been the reason behind the improved HR:NHR ratio (although it significantly improved the transformation efficiency).

I then tested deleting two other genes (*psbA*2 and *psb*28.2) using the one-step PCR-based method to check if the sequence of the *psbN* gene itself played a role in preventing consistent HR. However, no targeted gene knockouts were obtained for either of the two genes tested suggesting the integration is not influenced by the sequence of the gene targeted (Figure 4.18). Several examinations suggested that the sequence of km^R cassette used for selection is influencing the NHR events taking place since (i) PCR analysis showed that some of the PCR products of LT50.*psbN*::km^R and LT50.*psb28.2*::km^R were integrating in the same location as the naked km^R cassette seen in chapter 3, (ii) sequence analysis of a representative of each of the PCR products of

LT50.*psbN*::km^R and LT50.*psbA2*::km^R PCR products revealed the presence of nucleotide deletions, mutations and insertions near the junction of the km^R priming site and the homology extension when long tail primers were used, (iii) sequence analysis of the PCR product of the LT50.*psbN*::km^R exposed the presence of a stretch of 30 bp of the km^R cassette in the junction between the priming site and homology extension. All these findings suggest that using the km^R cassette in the one-step PCR-based method leads to nucleotide mutations, deletions and DNA shuffling that ultimately results in the creation of flanks that are not identified as flanks by the recipient strain and therefore lead to NHR events that target the PCR products to random locations or, in some cases, to the *speA/psbA2* locus.

In order to confirm these results, the km^R cassette was replaced with a *ble* selectable marker in creating the three gene (*psbN*, *psbA2* and *psb.28.2*) knockout products using the one-step PCR method, which were then transformed into Δ*recJ.*ble strain. Analysis of the putative transformants confirmed, for the first time, that the one-step PCR method was successful in creating targeted gene knockouts of the three genes in a good HR:NHR ratio. These results were very encouraging and further investigation into the possibility of generating the same gene knockouts using 30 bp homology showed that 50 bp was probably close to the minimal length of homology needed for this method using *ble* as the selectable marker. On the other hand, attempts to use the one-step PCR method to delete the phycobilisome operon (consisting of four genes) failed, suggesting that the one-step PCR method may not be feasible for large chromosomal deletions.

In conclusion, despite several reports on the failure to produce successful gene knockouts by means of double homologous recombination in *Synechocystis* with flanks less than 100 bp, this study established the possibility of creating knockouts using PCR products containing as little as 30 bp flanking regions. Nevertheless, the efficiency of the transformation did drop markedly with decreased flank sizes. Additionally, a method was developed that allowed for the creation of targeted gene knockouts in a single PCR step utilising the Δ*recJ.*ble strain as the recipient strain and *ble* as the selectable marker of choice.

Recommended next steps would be trying to create targeted gene knockouts using the developed one-step PCR method that would result in a phenotype

difference to be used as a quick screening method for gene knockouts. An example would be converting a sucrose sensitive strain of *Synechocystis* into a sucrose resistance strain. This work has already been initiated on a mutant strain of *Synechocystis* provided by David Lea Smith (Cambridge University, UK). In this strain, the *COX* gene, that encodes for a subunit of the thylakoid membrane-localised cytochrome *c* oxidase, was deleted by the insertion of a *npt*I.*sac*B cassette resulting in the creation of a strain that is sucrose sensitive (Lea-Smith et al., 2013) (Figure 7.1.A). The strain provided was then converted into a *recJ*- strain by deleting the *recJ* gene via the insertion of the *ble* cassette producing the Δcox::sacBnptI::Δ*recJ*::ble strain (Figure 7.1.B). The new strain created was both kanamycin and zeocin resistant but not sucrose resistant (Table 7.1). The next steps (currently on-going by Janet Waterhouse from Saul Purton's group) would be to generate PCR products with 50 bp homology to the *COX* gene and ~ 22 bp homologous to an *aadA* cassette (confers resistance to spectinomycin) and use the PCR product in a transformation to generate the Δcox::aadAΔrecJ::ble strain that would be sucrose and spectinomycin resistant (Figure 7.1.C).

Table 7.1: Expected antibiotic resistance patterns of the three strains.

Genotype	Resistance			
	Km	Zeo	Spec	suc
Δcox::sacBnptI (provided)	√	X	X	X
Δcox::sacBnptIΔrecJ::ble (created by myself)	√	√	X	X
Δcox::aadAΔrecJ::ble (future work)	X	√	√	√

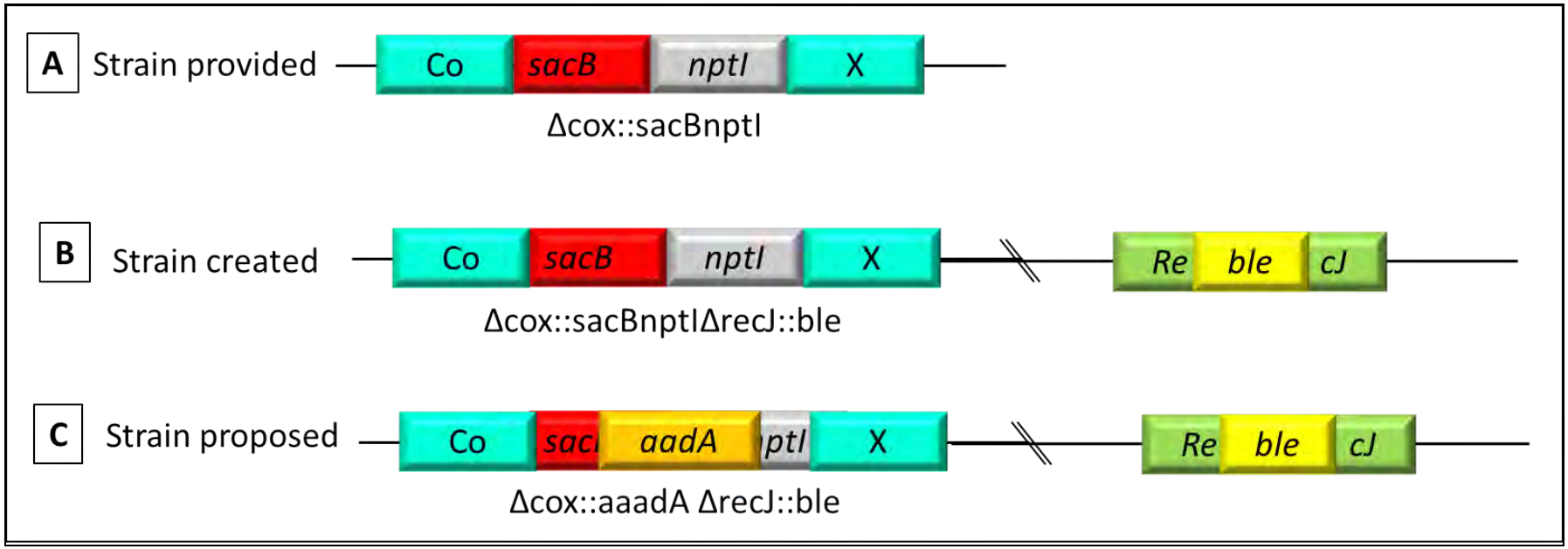


Figure 7.1: Schemetic representation of the Δcox::sacBnptI, the Δcox::sacBnptIΔrecJ::ble and the Δcox::aadAΔrecJ::ble strains.

(A)The Δcox::sacB.nptI strain provided by David Lea Smith (B)The Δcox::sacB.nptIΔrecJ::ble created in this work (C) On going work by Janet Waterhouse is being done to create the Δcox::aadAΔrecJ::ble strain.

7.1.3 The potential of engineering *Synechocystis* 6803 for photosynthetic terpene production

The production of terpenes in microbial hosts has gained a lot of attention in recent years due to the various industrial applications of this group of compounds (Sikkema et al., 1995, Dunlop et al., 2011, Ro et al., 2006). The commercial value of geraniol for example, lies in its application as an essential oil used in perfumes and cosmetics, and as a natural insect repellent. Furthermore, it holds potential as a fungible fuel or fuel precursor, while farnesene on the other hand, serves as a potential diesel replacement. One challenge facing the production of terpenes in a microbial host is the toxicity of the terpene on the host. Nevertheless, the production of geraniol has been achieved in both modified yeast (Oswald et al., 2007) and tomato (Davidovich-Rikanati et al., 2007) while farnesene has been produced in both genetically engineered *E. coli* (Wang et al., 2011) and *S. cerevisiae* (Peralta-Yahya et al., 2012, Brennan et al., 2012). However, to the best of my knowledge, the production of geraniol and farnesene has not been achieved in a photosynthetic microbial host.

Therefore, the aim of the work presented in chapter 5 was to explore the prospect of genetically engineering the isoprenoid biosynthetic pathway of *Synechocystis* for the photosynthetic production of geraniol and farnesene, to test the tolerance of *Synechocystis* to these terpenes, and to find ways to mitigate their toxic effects. This was done by expression of the synthase genes in *Synechocystis* under the control of two different promoters: a constitutive light-regulated promoter (*psbA2*) to be utilised in the expression of genes with products that are shown to be non-toxic to the cell, and a nickel inducible promoter (*nrsB*) that allows for the expression of toxic foreign genes. After confirming the functionality of the two developed expression vectors (pLAH.A2 and pLAH.nrsB), they were used in the attempted expression of both the geraniol synthase (GES) and α-farnesene synthase (FS) genes in *Synechocystis*.

The creation of a *Synechocystis* transformant line expressing the *GES* gene under the *psbA2* promoter was not possible as work had to be terminated since no *E. coli* transformants containing the pLAH.A2.GES plasmid were obtained possibly due to the active expression of the toxic *GES* gene in *E. coli* under the *psbA2* promoter (which is recognised in *E. coli*) resulting in cell death. This conclusion is perfectly viable since (i) the antimicrobial effects of geraniol have

been previously reported by several researchers (Andoğan et al., 2002, Sato et al., 2007) and (ii) its highly toxic effect on *Synechocystis* was demonstrated in work presented in chapter 5. On the other hand, despite obtaining *E. coli* transformants for the pLAH.nrsB.GES plasmid, transformation of *Synechocystis* failed to generate true transformants. It is possible that the *nrsB* promoter showed a "leaky" expression of the gene that caused cell death. *E. coli* cells were not affected since there are no reports of the *nrsB* promoter being active in *E. coli*. Unfortunately, since no *Synechocystis* transformants were obtained, no further work related to expressing *GES* in *Synechocystis* was feasible.

At the same time as the molecular work was taking place in our lab, engine test were carried out by Paul Hellier at the Department of Mechanical Engineering at UCL and results confirmed that geraniol is not suitable as either a diesel or gasoline replacement fuel as its combustion characteristics are not close to fossil fuel and therefore cannot be considered a "drop in fuel". However, engine tests of other terpenes showed that farnesene and geranial (citral-a) were better diesel replacement fuels, while linalool and citronellene were good gasoline replacements. The engine results prompted interest in testing the feasibility of expressing the five terpenes in *Synechocystis*. By examining the isoprenoid pathway, together with results from toxicity tests conducted on *Synechocystis* using these five selected terpenes indicated that farnesene was the most promising terpene that could potentially and conveniently be produced in a controlled manner in *Synechocystis* due to (i) its reasonably low toxicity (ii) the fact that it only requires the introduction of a single gene and (iii) the co-inoculation of high amounts of farnesene (1%) in the presence of a media additive or a "two-phase" culture system shows potential for the production of reasonably high amounts of this terpene in *Synechocystis*.

Transformants were readily obtained from both the pLAH.A2 and the pLAH.nrsB plasmids carrying the codon optimised α-farnesene synthase gene and western blot analysis confirmed the successful expression of the synthase under both promoters. In order to achieve higher levels of the α-farnesene synthase under the *psbA2* promoter, cultures need to be incubated under low light levels (20 μmol/m^2/s) until they reach to an OD_{750} of ~0.7 followed by a shift to high light (200 μmol/m^2/s). While, for optimum expression of the synthase under the *nrsB* promoter, cultures need to be induced at an OD_{750} (0.5 - 0.7) for 7.5 hours. Contrary to expectations, the actual concentration of the nickel inducer (in the

range tested) didn't seem to have an effect on the level of expression. This is contrary to what was observed with the expression of the *alkL* gene under the same promoter in chapter 6 where more of the protein was observed with increased nickel concentration. It might be possible that the steady-state level of the FS protein is primarily dictated by rates of degradation rather than rates of synthesis.

After successfully demonstrating the expression of α-farnesene synthase under both promoters, the next step was to try to detect farnesene production in the strains. Unfortunately, all attempts to detect the product using GC-MS were unsuccessful even after testing various extractions methods. Despite the inability to detect farnesene, comparative growth studies conducted on the strain expressing the synthase under the *nrsB* promoter and that of a negative control showed that the synthase-expressing strain appeared phenotypically different from the control post nickel induction. The fact that the *Synechocystis* culture expressing the synthase only showed deteriorated growth "post" nickel induction is a clear indication that the expressed synthase is adding a metabolic burden in the cell, possibly as a consequence of substrate depletion (IPP and DMAPP depletion), a common observation with research aimed at expressing terpenes in microbes (Oswald et al., 2007, Glick, 1995, Goff and Goldberg, 1985, Martin et al., 2003).

To summarize, the expression of the *GES* gene in *Synechocystis* proved very challenging due to the highly toxic effect of geraniol on the cell whereas the α-farnesene synthase gene was successfully expressed in *Synechocystis* under both the *psbA2* and *nrsB* promoters. However, no farnesene was detected from the synthase expressing strains possibly as a consequence of substrate deficiency. When working with biological systems, one may expect to encounter serious difficulties when it comes to designing optimization strategies for robust cellular performance (Angermayr et al., 2009). I believe that the recommended next steps that need to be taken in order to ultimately produce detectable amounts of farnesene in the farnesene synthase strains created in this work are to: (1) test the activity of the α-farnesene synthase by performing standard enzyme activity assays; (2) exogenously adding the rate limiting substrates (IDD and DMAPP) and check for the accumulation and detection of farnesene (3) heterologously express the yeast mevalonate pathway to increase the availability of the substrates that will ultimately lead to an increased titer of farnesene (similar

to work done by Wang et al. (2011), in engineered *E. coli* strains). Furthermore, since two reliable expression vectors have been developed in this work, it is worth testing the expression of a range of foreign genes in *Synechocystis,* particularly gene products with pharmaceutical applications especially after the successful expression (by several members of our group) of a range of proteins in *Synechocystis* using the pLAH.A2 and pLAH.nrsB expression vectors (Figure 7.2). Furthermore, the application of the two expression vectors can be extended to express an operon of genes under the control of the *nrsB* promoter. Ben Mackrow (a previous MSc student in the group) used Gibson assembly to build a pLAH.A2 based plasmid expressing an operon of four genes encoding three antibiotics and a green florescent protein (GFP) (Figure 7.2.A). The ability to express an operon under the control of the *nrsB* promoter would allow for wider studies aimed at the controlled expression of multigenic metabolic pathways in this model organism.

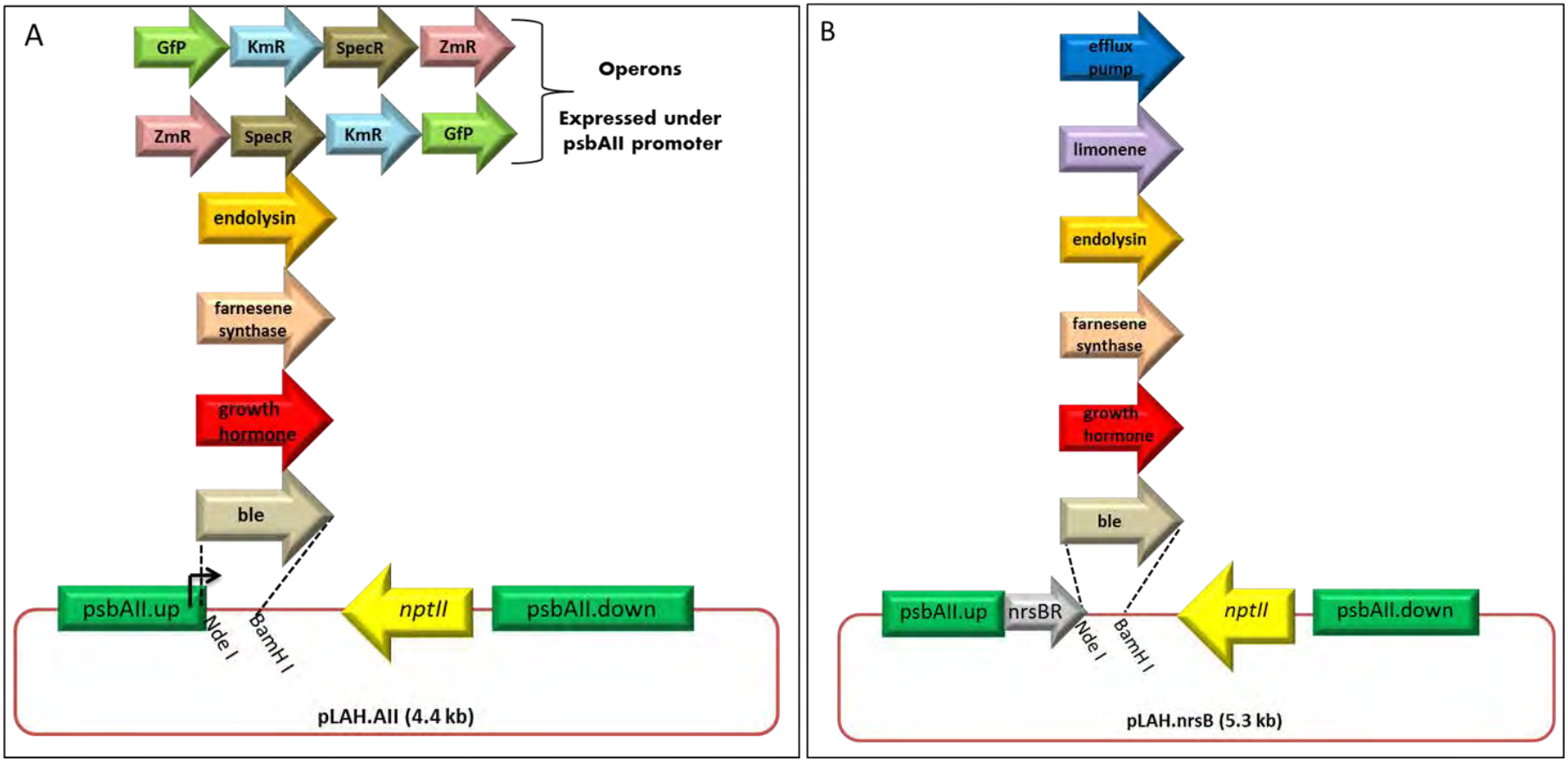


Figure 7.2 : The expression of various proteins in the pLAH.A2 and pLAH.nrsB expression vectors

(A) Proteins and operons expressed under the constitutive (B) Proteins expressed under the constitutive *nrsB* promoter. **Note**: The Ble, farnesene synthase and efflux pump proteins were expressed by myself, Endolysin by Laura Stoffels, Human Growth Hormone by Janet Waterhouse, Limonene synthase by Alice Liu and the operons by Ben Mackrow. **Key:** GfP: green fluorescent protein, $Spec^R$; spectinomycin resistance cassette, km^R; kanamycin resistance cassette, Zm^R; zeomycin resistance cassette.

7.1.4 The expression of AlkL in *Synechocystis* to facilitate substrate delivery and product removal of hydrocarbons

Until recently, the function of the AlkL protein present in *Pseudomonas putida* GPo1 was not clearly identified with most researchers hypothesizing its role to be involved in the transport and degradation of alkanes (van Beilen et al., 2001, Hong et al., 2006, Hearn et al., 2009, van Beilen et al., 1992). Recently, this hypothesis was confirmed after the functional expression of the AlkL protein was shown to enable over 120 fold improvement of uptake and conversion of C_{12}-C_{16} alkanes by an intracellular enzyme expressed in *E. coli* strain (Grant, 2012). Further work by Grant and colleagues also showed that alkL could lead to up to two fold higher yields of C_{13}-C_{16} alkanes produced by an *E. coli* strain engineered to produce alkanes. My aim in work carried out in chapter 6 was to (i) express the AlkL transporter in the natural alkane producing *Synechocystis* under different promoters as a demonstration of the expression of a heterologous outer membrane protein, (ii) test if the increase in alkane yields observed in *E. coli* would also be seen in *Synechocystis,* (iii) test for any changes to the ratio of secreted and intracellular accumulation of naturally produced alkanes in *Synechocystis,* and finally (iv) test the effect of a two-phase culture system and detergent on the extraction of intracellular alkanes, which had been shown to enhance alkane yields in *E. coli*. Eventually, all these parameters were to be compared to results obtained in similar studies on *E. coli* conducted by Christopher Grant.

My studies on *Synechocystis* showed that the expression of *alkL* under the *nrsB* promoter was successful and the protein accumulation was regulated, in a dose-dependent manner, using different concentrations of the inducer. This titratable expression of the AlkL protein is essential in order to avoid over expression of the protein that may lead to cell toxicity (Grant, 2012). On the other hand, the transformation of *E. coli* with the pLAH.A2.alkL expression vector did not result in any transformants; suggesting that the functional expression of the toxic gene under the *psbA2* promoter in *E. coli* resulted in cell death. This is similar to results obtained in chapter 5 where attempts to express the toxic *GES* gene in *E. coli* under the same promoter failed. We therefore concluded that achieving functional expression of genes for toxic products under the *psbA2* promoter in both *E. coli*

and *Synechocystis* is not feasible and therefore no further work expressing the protein transporter under the *psbA2* promoter was carried out.

However, further analysis using the *nrsB* promoter showed that the AlkL protein is most likely being expressed in a functional form indicated by a clear phenotypic change to the strain indicated by the loss of pigmentation compared to the negative control. Additionally, the dry cell weight of the *alkL* expressing strain was found to be inversely proportional to the inducer concentration. Both these findings are in accordance to what was observed with *E. coli* and can be used as an indication of the functional expression of the protein in *Synechocystis*. However, GC results conducted to examine the role of the AlkL protein in accumulating intracellular levels of alkanes in *Synechocystis* showed a general reduction in the intracellular levels of alkanes upon AlkL expression. However, this result is not conclusive since the same result was observed with negative controls (lacking the AlkL transporter) indicating that the reduction in intracellular levels of the naturally produced alkanes is most probably attributed to prolonged nickel induction rather than the expression of the AlkL protein. Nonetheless, the results on *Synechocystis* were contrary to what was shown in *E. coli* cells in which case the expression of the AlkL protein resulted in the increase in the intracellular yields of alkanes by two fold (1.5 for heptadecane). Moreover, the GC analysis of the supernatant and pellet fractions showed no evidence of export of the C_{17} leading to the conclusion that the AlkL protein did not result in the transport of heptadecane in *Synechocystis*. This is an intriguing result in itself as it may be that the heptadecane is being retained by the cell and therefore the natural function of heptadecane in *Synechocystis*, which is currently not identified, is likely to be within the cell (such as carbon storage) rather than for an export related function (such as signaling).

On the other hand, a larger fraction of the C_{17} alkane was found in the supernatant of cultures grown in a two-phase culture system compared to those grown in a single phase indicating the possible role of the added n-dodecane in “milking out” the naturally produced C_{17} alkane. In *E. coli*, the addition of the detergent and n-dodecane led to a much higher extraction with a shift from 20% alkane recovery from the medium (with no Triton X-100 or n-dodecane) to up to 90% recovery of the alkane (Christopher Grant, personal communication). It also appears that the addition of n-dodecane second phase has a positive effect on

the cell since cultures grown in the presence of n-dodecane displayed a darker green appearance possibility as a result of increased chlorophyll content of the cells. Nevertheless, the role of n-dodecane in promoting growth is still not clear although the cell density measurements did not show a positive correlation with the presence of n-dodecane indicating that the solvent was not necessary improving growth.

In conclusion, the AlkL protein was successfully expressed in *Synechocystis* in a dose-dependent manner under the control of the *nrsB* inducible promoter. The reduction of dry cell weight and the clear phenotype change of the AlkL expressing strains are strong indications on the functionality or partial activity of the transporter. However, the role of the AlkL protein in the transport and accumulation of intracellular levels of alkanes is still not conclusive. Nevertheless, results presented in this chapter provide a foundation for a deeper understanding and further development of the AlkL bio-pump in *Synechocystis* and also has implications for understanding the natural function of heptadecane in *Synechocystis*. Hypothesis such as the following can be investigated: (i) the inner membrane of *Synechocystis* may be the barrier to alkane transport, (ii) there are natural alkane importer proteins already expressed in *Synechocystis* which prevent loss of heptadecane through diffusion across the membrane, (iii) or the possibility of the heptadecane being compartmentalised in the cell to serve a specific function. For this, the following sets of experiments are recommended:

- **Functional expression**: For clear evidence on the functionality of the AlkL protein, *alkL* expressing strains can be incubated with certain dyes such aspropidium iodide, fluorescein or Nile red, and monitored for uptake using flow cytometry.

- **Translocation of the protein to the outer membrane**: In order to confirm that the AlkL protein was indeed expressed in the outer membrane fraction, the cell can be fractionated and a western blot performed on the outer membrane fraction.

- **C_{12} Alkane analysis:** the GC analysis of the C_{12} dodecane in the pellet and supernatant fractions both pre and post *alkL* induction in the presence and absence of the detergent should be conducted alongside negative controls for a better understanding of the behavior of this auxiliary phase on *Synechocystis* cells. This would help answer questions such as: is it possible that n-dodecane is naturally taken up by *Synechocystis* cells? How much does the detergent enhance n-dodecane uptake? And finally, would the expression of *alkL* affect the transport of the alkane into the cell?

- **C_{17} Alkane analysis:** The GC analysis of the C_{17} should be repeated including more controls to obtain a clear answer on the role the AlkL protein has on the intracellular levels of heptadecane.

7.2 Conclusions, challenges and perspectives

There has been recent interest in exploring cyanobacteria as a platform for the production of chemicals and biofuels from CO_2 and sunlight (Zhou and Li, 2010, Yu et al., 2013, Hu et al., 2013). Although a lot of metabolic engineering tools have been developed for the genetically amendable model organism *Synechocystis* 6803 and a range of chemicals, health products, and potential biofuels have already been reported (chapter 1), the production rates and titers of the products in engineered *Synechocystis* are currently very low and cannot compete with heterotrophic fermentation. One example is the production rate of fatty acids in engineered *E. coli* which is reported to be 5.2 g/L in three days (Zhang et al., 2012) while engineering the same pathway in *Synechocystis* yielded less than 0.2 g/L after two weeks of cultivation (Liu et al., 2011).

Furthermore, most of the products synthesised in *Synechocystis* and other cyanobacteria involved the expression of short pathways (containing 2–5 heterologous enzymes). It is apparent that as the heterologous pathway gets larger, product synthesis becomes more complicated and difficult to achieve (Yu et al., 2013). Therefore, the synthesis of products in *Synechocystis* needs to be currently focused on the expression of less complicated pathways until a more advanced synthetic biology toolbox is developed for this cyanobacterium. The quick one-step PCR-based method for creating targeted gene knockouts/knock-

ins in *Synechocystis* developed here and presented in chapter 4 will aid in speeding up molecular work. On the other hand, it is common that even the expression of a single gene (*e.g.* α-farnesene synthase seen in chapter 5) in *Synechocystis* may lead to the deterioration of the engineered strain. Nonetheless, a range of heterologous genes and proteins were successfully expressed in our lab using the *psbA2* and *nrsB* native promoters as shown in chapter 5, 6 and 7.

In order to take photosynthetic bio-production in *Synechocystis* to an industrial scale, an interdisciplinary approach needs to be taken, similar to what was done with work presented in chapters 5 and 6 of this thesis, that involves microbiologists, synthetic biologists and chemical engineers to create a synthetic toolbox for *Synechocystis* similar to that developed for model heterotrophic microbes (*E. coli* and *S. cerviseae*) in order to solve issues such as the low efficiency of cell metabolism, optimizing growth conditions, limitations in bioprocesses, effect of product toxicity and cost associated with product recovery.

References

ADAMS, D. 2002. Cyanobacteria in symbiosis with hornworts and liverworts. *Cyanobacteria in symbiosis.* Dordrech: Kluwer Academic Publishers.
ADAMS, D. G. & DUGGAN, P. S. 2008. Cyanobacteria–bryophyte symbioses. *Journal of Experimental Botany,* 59**,** 1047-1058.
ÅHMAN, M. 2010. Biomethane in the transport sector—An appraisal of the forgotten option. *Energy Policy,* 38**,** 208-217.
AKIYAMA, Y. 2009. Quality Control of Cytoplasmic Membrane Proteins in Escherichia coli. *Journal of Biochemistry,* 146**,** 449-454.
AL-HAJ, L. 2012. Biofuels from Microorganisms. *APBN,* 16**,** 30-33.
ALPER, H. & STEPHANOPOULOS, G. 2009. Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nat Rev Micro,* 7**,** 715-723.
ANDOĞAN, B., BAYDAR, H., KAYA, S., DEMIRCI, M., ÖZBAŞAR, D. & MUMCU, E. 2002. Antimicrobial activity and chemical composition of some essential oils. *Archives of Pharmacal Research,* 25**,** 860-864.
ANDRIANANTOANDRO, E., BASU, S., KARIG, D. K. & WEISS, R. 2006. Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol,* 2**,** 2006.0028.
ANGERMAYR, S. A., HELLINGWERF, K. J., LINDBLAD, P. & TEIXEIRA DE MATTOS, M. J. 2009. Energy biotechnology with cyanobacteria. *Current Opinion in Biotechnology,* 20**,** 257-263.
ANTONI, D., ZVERLOV, V. V. & SCHWARZ, W. H. 2007. Biofuels from microbes. *Applied Microbiology and Biotechnology,* 77**,** 23-35.
ASTIER, C., ELMORJANI, K., MEYER, I., JOSET, F. & HERDMAN, M. 1984. Photosynthetic mutants of the cyanobacteria Synechocystis sp. strains PCC 6714 and PCC 6803: sodium p-hydroxymercuribenzoate as a selective agent. *Journal of Bacteriology,* 158**,** 659-664.
ATSUMI, S., CANN, A. F., CONNOR, M. R., SHEN, C. R., SMITH, K. M., BRYNILDSEN, M. P., CHOU, K. J., HANAI, T. & LIAO, J. C. 2008a. Metabolic engineering of Escherichia coli for 1-butanol production. *Metab Eng,* 10**,** 305-11.
ATSUMI, S., HANAI, T. & LIAO, J. C. 2008b. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature,* 451**,** 86-89.
ATSUMI, S., HIGASHIDE, W. & LIAO, J. C. 2009. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nat Biotech,* 27**,** 1177.
AZEVEDO, M. M. B., PEREIRA, A. Q., CHAVES, F. C. M., BIZZO, H. R., ALVIANO, C. S. & ALVIANO, D. S. 2012. Antimicrobial activity of the essential oils from the leaves of two morphotypes of Croton cajucara Benth. *Journal of Essential Oil Research,* 24**,** 351-357.
BARNWAL, B. K. & SHARMA, M. P. 2005. Prospects of biodiesel production from vegetable oils in India. *Renewable and Sustainable Energy Reviews,* 9**,** 363-378.
BARTEN, R. & LILL, H. 1995. DNA-uptake in the naturally competent cyanobacterium, synechocystis sp pcc-6803. *FEMS Microbiology Letters,* 129**,** 83-88.
BEER, L. L., BOYD, E. S., PETERS, J. W. & POSEWITZ, M. C. 2009. Engineering algae for biohydrogen and biofuel production. *Current Opinion in Biotechnology,* 20**,** 264.

BELAY, A., OTA, Y., MIYAKAWA, K. & SHIMAMATSU, H. 1993. Current knowledge on potential health benefits of Spirulina. *Journal of Applied Phycology,* 5**,** 235-241.
BELOIN, C., DEIGHAN, P., DOYLE, M. & DORMAN, C. J. 2003. Shigella flexneri 2a strain 2457T expresses three members of the H-NS-like protein family: characterization of the Sfh protein. *Mol Genet Genomics,* 270**,** 66 - 77.
BERGEY, D. H. & HOLT, J. G. 1994. *Bergey's manual of determinative bacteriology,* Baltimore, Williams & Wilkins.
BERLA, B. M., SAHA, R., IMMETHUN, C. M., MARANAS, C. D., MOON, T. S. & PAKRASI, H. 2013. Synthetic Biology of Cyanobacteria: Unique Challenges and Opportunities. *Frontiers in Microbiology,* 4.
BERTANI, G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J Bacteriol,* 62**,** 293-300.
BIRNBOIM, H. C. & DOLY, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res,* 7**,** 1513-23.
BISHOP, C. 2002. *An investigation of the small polypeptides of photosystem II.* PhD thesis, University College London.
BRENNAN, T. C. R., TURNER, C. D., KRÖMER, J. O. & NIELSEN, L. K. 2012. Alleviating monoterpene toxicity using a two-phase extractive fermentation for the bioproduction of jet fuel mixtures in Saccharomyces cerevisiae. *Biotechnology and Bioengineering,* 109**,** 2513-2522.
BROCKS, J. J., LOGAN, G. A., BUICK, R. & SUMMONS, R. E. 1999. Archean molecular fossils and the early rise of eukaryotes. *Science,* 285**,** 1033-6.
CARTER, O. A., PETERS, R. J. & CROTEAU, R. 2003. Monoterpene biosynthesis pathway construction in Escherichia coli. *Phytochemistry,* 64**,** 425-33.
CASTENHOLZ, R. W. 1988. [3] Culturing methods for cyanobacteria. *In:* LESTER PACKER, A. N. G. (ed.) *Methods in Enzymology.* Academic Press.
CHAKRABARTY, A. M., CHOU, G. & GUNSALUS, I. C. 1973. Genetic regulation of octane dissimilation plasmid in Pseudomonas. *Proc Natl Acad Sci U S A,* 70**,** 1137-40.
CHAMBON, C., LADEVEZE, V., OULMOUDEN, A., SERVOUSE, M. & KARST, F. 1990. Isolation and properties of yeast mutants affected in farnesyl diphosphate synthetase. *Current Genetics,* 18**,** 41-46.
CHAMBON, C., LADEVEZE, V., SERVOUSE, M., BLANCHARD, L., JAVELOF, C., VLADESCU, B. & KARST, F. 1991. Sterol pathway in yeast. Identification and properties of mutant strains defective in mevalonate diphosphate decarboxylase and farnesyl diphosphate synthetase. *Lipids,* 26**,** 633-636.
CHANG, M. C. Y., EACHUS, R. A., TRIEU, W., RO, D.-K. & KEASLING, J. D. 2007. Engineering Escherichia coli for production of functionalized terpenoids using plant P450s. *Nat Chem Biol,* 3**,** 274-277.
CHAUVAT, F., ROUET, P., BOTTIN, H. & BOUSSAC, A. 1989. Mutagenesis by random cloning of an Escherichia coli kanamycin resistance gene into the genome of the cyanobacterium Synechocystis PCC 6803: Selection of mutants defective in photosynthesis. *Molecular and General Genetics MGG,* 216**,** 51-59.

CHAUVAT, F., VRIES, L., ENDE, A. & ARKEL, G. A. 1986. A host-vector system for gene cloning in the cyanobacterium Synechocystis PCC 6803. *Molecular and General Genetics MGG,* 204**,** 185-191.
CHAVEROCHE, M. K., GHIGO, J. M. & D'ENFERT, C. 2000. A rapid method for efficient gene replacement in the filamentous fungus Aspergillus nidulans. *Nucleic Acids Res,* 28**,** E97.
CHEAH, Y. E., ALBERS, S. C. & PEEBLES, C. A. 2013. A novel counter-selection method for markerless genetic modification in Synechocystis sp. PCC 6803. *Biotechnol Prog,* 29**,** 23-30.
CHEESBROUGH, T. M. & KOLATTUKUDY, P. E. 1984. Alkane biosynthesis by decarbonylation of aldehydes catalyzed by a particulate preparation from Pisum sativum. *Proc Natl Acad Sci U S A,* 81**,** 6613-7.
CHIANG, G., SCHAEFER, M. & GROSSMAN, A. 1992. Transformation of the filamentous cyanobacterium Fremyella diplosiphon by conjugaison or electroporation. *Plant physiology and biochemistry,* 30**,** 315-325.
CHISTI, Y. 2007. Biodiesel from microalgae. *Biotechnology Advances,* 25**,** 294-306.
CHU, H.-A., NGUYEN, A. P. & DEBUS, R. J. 1994. Site-Directed Photosystem II Mutants with Perturbed Oxygen-Evolving Properties. 2. Increased Binding or Photooxidation of Manganese in the Absence of the Extrinsic 33-kDa Polypeptide In vivo. *Biochemistry,* 33**,** 6150-6157.
CIFERRI, O. & TIBONI, O. 1985. The Biochemistry and Industrial Potential of Spirulina. *Annual Review of Microbiology,* 39**,** 503-526.
COGNE, G., CORNET, J. F. & GROS, J. B. 2005. Design, operation, and modeling of a membrane photobioreactor to study the growth of the Cyanobacterium Arthrospira platensis in space conditions. *Biotechnol Prog,* 21**,** 741-50.
CONNOR, M. R. & ATSUMI, S. 2010. Synthetic Biology Guides Biofuel Production. *Journal of Biomedicine and Biotechnology,* 2010.
CORDOBA, E., SALMI, M. & LEON, P. 2009. Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *Journal of Experimental Botany,* 60**,** 2933-2943.
DATSENKO, K. A. & WANNER, B. L. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A,* 97**,** 6640 - 6645.
DAUM, M., HERRMANN, S., WILKINSON, B. & BECHTHOLD, A. 2009. Genes and enzymes involved in bacterial isoprenoid biosynthesis. *Current Opinion in Chemical Biology,* 13**,** 180-188.
DAVIDOVICH-RIKANATI, R., SITRIT, Y., TADMOR, Y., IIJIMA, Y., BILENKO, N., BAR, E., CARMONA, B., FALLIK, E., DUDAI, N., SIMON, J. E., PICHERSKY, E. & LEWINSOHN, E. 2007. Enrichment of tomato flavor by diversion of the early plastidial terpenoid pathway. *Nature Biotechnology,* 25**,** 899-901.
DE TARSAC, N. & HOUMARD, J. 1987. Advances in cyanobacterial molecular genetics. *The Cyanobacteria.***,** 251-302.
DEDONDER, R. 1966. [86] Levansucrase from Bacillus subtilis. *In:* ELIZABETH F. NEUFELD, V. G. (ed.) *Methods in Enzymology.* Academic Press.

DEN HAAN, R., ROSE, S. H., LYND, L. R. & VAN ZYL, W. H. 2007. Hydrolysis and fermentation of amorphous cellulose by recombinant Saccharomyces cerevisiae. *Metab Eng,* 9**,** 87-94.
DENG, M. D. & COLEMAN, J. R. 1999. Ethanol Synthesis by Genetic Engineering in Cyanobacteria. *Applied and Environmental Microbiology,* 65**,** 523-528.
DENNIS, M. & KOLATTUKUDY, P. E. 1992. A cobalt-porphyrin enzyme converts a fatty aldehyde to a hydrocarbon and CO. *Proceedings of the National Academy of Sciences,* 89**,** 5306-5310.
DERBISE, A., LESIC, B., DACHEUX, D., GHIGO, J. M. & CARNIEL, E. 2003. A rapid and simple method for inactivating chromosomal genes in Yersinia. *FEMS Immunol Med Microbiol,* 38**,** 113 - 116.
DESIKACHARY, T. 1959. Cyanophyta. New Delhi: Indian Council of Agricultural Research.
DEWICK, P. M. 2002. The biosynthesis of C5-C25 terpenoid compounds. *Nat Prod Rep,* 19**,** 181-222.
DEXTER, J. & FU, P. 2009. Metabolic engineering of cyanobacteria for ethanol production. *Energ Environ Sci,* 2**,** 857 - 864.
DISMUKES, G. C., CARRIERI, D., BENNETTE, N., ANANYEV, G. M. & POSEWITZ, M. C. 2008. Aquatic phototrophs: efficient alternatives to land-based crops for biofuels. *Current Opinion in Biotechnology,* 19**,** 235-240.
DUCAT, D. C., SACHDEVA, G. & SILVER, P. A. 2011a. Rewiring hydrogenase-dependent redox circuits in cyanobacteria. *Proceedings of the National Academy of Sciences,* 108**,** 3941-3946.
DUCAT, D. C., WAY, J. C. & SILVER, P. A. 2011b. Engineering cyanobacteria to generate high-value products. *Trends in Biotechnology,* 29**,** 95-103.
DUNLOP, M. 2011. Engineering microbes for tolerance to next-generation biofuels. *Biotechnology for Biofuels,* 4**,** 1-9.
DUNLOP, M. J., DOSSANI, Z. Y., SZMIDT, H. L., CHU, H. C., LEE, T. S., KEASLING, J. D., HADI, M. Z. & MUKHOPADHYAY, A. 2011. Engineering microbial biofuel tolerance and export using efflux pumps. *Mol Syst Biol,* 7.
DZELZKALNS, V. A. & BOGORAD, L. 1986. Stable transformation of the cyanobacterium synechocystis sp pcc-6803 induced by uv irradiation. *Journal of Bacteriology,* 165**,** 964-971.
EATON-RYE, J. J. 2004. The Construction of Gene Knockouts in the Cyanobacterium Synechocystis sp. PCC 6803. *Photosynthesis Research Protocols.*
EGGINK, G., VAN LELYVELD, P. H., ARNBERG, A., ARFMAN, N., WITTEVEEN, C. & WITHOLT, B. 1987. Structure of the Pseudomonas putida alkBAC operon. Identification of transcription and translation products. *J Biol Chem,* 262**,** 6400-6.
EISA. 2007. *U.S. Congressional Research Service. Energy Independence and Security Act of 2007: A Summary of Major Provisions* [Online]. Available: http://energy.senate.gov/public/_files/RL342941.pdf.
ELHAI, J. 1994. Genetic techniques appropriate for the biotechnological exploitation of cyanobacteria. *Journal of Applied Phycology,* 6**,** 177-186.

ELHAI, J. & WOLK, C. P. 1988. [83] Conjugal transfer of DNA to cyanobacteria. *In:* LESTER PACKER, A. N. G. (ed.) *Methods in Enzymology.* Academic Press.
ENGLER, C. & MARILLONNET, S. 2011. Generation of families of construct variants using golden gate shuffling. *Methods Mol Biol,* 729**,** 167-81.
ERMAKOVA-GERDES, S. & VERMAAS, W. 1999. Development of A psbA-less/psbD-less Strain of Synechocystis sp. PCC 6803 for Simultaneous Mutagenesis of the D1 and D2 Proteins of Photosystem II. *In:* PESCHEK, G., LÖFFELHARDT, W. & SCHMETTERER, G. (eds.) *The Phototrophic Prokaryotes.* Springer US.
ERSHOV, Y. V., GANTT, R. R., CUNNINGHAM JR, F. X., JR. & GANTT, E. 2002. Isoprenoid biosynthesis in Synechocystis sp. strain PCC6803 is stimulated by compounds of the pentose phosphate cycle but not by pyruvate or deoxyxylulose-5-phosphate. *J Bacteriol,* 184**,** 5045-51.
FELIZARDO, P., NEIVA CORREIA, M. J., RAPOSO, I., MENDES, J. F., BERKEMEIER, R. & BORDADO, J. M. 2006. Production of biodiesel from waste frying oils. *Waste Management,* 26**,** 487-494.
FORTMAN, J. 2008. Biofuel alternatives to ethanol: pumping the microbial well. *Trends Biotechnol,* 26**,** 375 - 81.
FORTMAN, J. L., CHHABRA, S., MUKHOPADHYAY, A., CHOU, H., LEE, T. S., STEEN, E. & KEASLING, J. D. 2008. Biofuel alternatives to ethanol: pumping the microbial well. *Trends Biotechnol,* 26**,** 375-81.
FU, P. 2009. Genome-scale modeling of Synechocystis sp. PCC 6803 and prediction of pathway insertion. *Journal of Chemical Technology & Biotechnology,* 84**,** 473-483.
GAO, K. 1998. Chinese studies on the edible blue-green alga, Nostoc flagelliforme: a review. *Journal of Applied Phycology,* 10**,** 37-49.
GARCÍA-DOMÍNGUEZ, M., LOPEZ-MAURY, L., FLORENCIO, F. J. & REYES, J. C. 2000. A Gene Cluster Involved in Metal Homeostasis in the Cyanobacterium Synechocystis sp. Strain PCC 6803. *Journal of Bacteriology,* 182**,** 1507-1514.
GEU-FLORES, F., NOUR-ELDIN, H. H., NIELSEN, M. T. & HALKIER, B. A. 2007. USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic acids research,* 35**,** e55.
GIBSON, D. G., YOUNG, L., CHUANG, R.-Y., VENTER, J. C., HUTCHISON, C. A. & SMITH, H. O. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Meth,* 6**,** 343-345.
GILPIN, S., HUI, X. & MAIBACH, H. 2010. In vitro human skin penetration of geraniol and citronellol. *Dermatitis,* 21**,** 41-8.
GLICK, B. R. 1995. Metabolic load and heterologous gene expression. *Biotechnol Adv,* 13**,** 247-61.
GOFF, S. A. & GOLDBERG, A. L. 1985. Production of abnormal proteins in E. coli stimulates transcription of lon and other heat shock genes. *Cell,* 41**,** 587-95.
GOLDEMBERG, J. 2007. Ethanol for a Sustainable Energy Future. *Science,* 315**,** 808-810.

GOLDEN, J. W. & WIEST, D. R. 1988. Genome rearrangement and nitrogen fixation in Anabaena blocked by inactivation of xisA gene. *Science,* 242**,** 1421-3.
GOLDEN, S. S. 1988. [80] Mutagenesis of cyanobacteria by classical and gene-transfer-based methods. *In:* LESTER PACKER, A. N. G. (ed.) *Methods in Enzymology.* Academic Press.
GOLDEN, S. S., BRUSSLAN, J. & HASELKORN, R. 1987. Genetic engineering of the cyanobacterial chromosome. *Methods Enzymol,* 153**,** 215-31.
GOLDEN, S. S. & SHERMAN, L. A. 1984. Optimal conditions for genetic transformation of the cyanobacterium Anacystis nidulans R2. *Journal of Bacteriology,* 158**,** 36-42.
GONZÁLEZ-FERNÁNDEZ, C., SIALVE, B., BERNET, N. & STEYER, J.-P. 2012. Impact of microalgae characteristics on their conversion to biofuel. Part II: Focus on biomethane production. *Biofuels, Bioproducts and Biorefining,* 6**,** 205-218.
GRANT, C. 2012. *Evaluation of the Bio-oxidation of Alkanes.* EngD Thesis, University College London.
GRANT, C. 2013. Bio-ISPR transporter plug-ins for enhancing microbial bioalkane synthesis (un-published).
GRANT, C., PINTO, A. C., LUI, H. P., WOODLEY, J. M. & BAGANZ, F. 2012. Tools for characterizing the whole-cell bio-oxidation of alkanes at microscale. *Biotechnol Bioeng,* 109**,** 2179-89.
GRIGORIEVA, G. & SHESTAKOV, S. 1982. Transformation in the cyanobacterium Synechocystis sp. 6803. *FEMS Microbiology Letters,* 13**,** 367-370.
GUSTAFSSON, C., GOVINDARAJAN, S. & MINSHULL, J. 2004. Codon bias and heterologous protein expression. *Trends Biotechnol,* 22**,** 346-53.
HAN, J., MCCARTHY, E., HOEVEN, W., CALVIN, M. & BRADLEY, W. 1968. Organic geochemical studies, ii. A preliminary report on the distribution of aliphatic hydrocarbons in algae, in bacteria, and in a recent lake sediment. *Proc Natl Acad Sci U S A,* 59**,** 29 - 33.
HANAI, T., ATSUMI, S. & LIAO, J. C. 2007. Engineered Synthetic Pathway for Isopropanol Production in Escherichia coli. *Applied and Environmental Microbiology,* 73**,** 7814-7818.
HANKAMER, B., LEHR, F., RUPPRECHT, J., MUSSGNUG, J. H., POSTEN, C. & KRUSE, O. 2007. Photosynthetic biomass and H2 production by green algae: from bioengineering to bioreactor scale-up. *Physiol Plant,* 131**,** 10-21.
HANSEL, A., PATTUS, F., JURGENS, U. J. & TADROS, M. H. 1998. Cloning and characterization of the genes coding for two porins in the unicellular cyanobacterium Synechococcus PCC 6301. *Biochim Biophys Acta,* 1399**,** 31-9.
HE, Q., SCHLICH, T., PAULSEN, H. & VERMAAS, W. 1999. Expression of a higher plant light-harvesting chlorophyll a/b-binding protein in Synechocystis sp. PCC 6803. *Eur J Biochem,* 263**,** 561-70.
HEARN, E. M., PATEL, D. R., LEPORE, B. W., INDIC, M. & VAN DEN BERG, B. 2009. Transmembrane passage of hydrophobic compounds through a protein channel wall. *Nature,* 458**,** 367-70.

HEIDORN, T., CAMSUND, D., HUANG, H. H., LINDBERG, P., OLIVEIRA, P., STENSJO, K. & LINDBLAD, P. 2011. Synthetic biology in cyanobacteria engineering and analyzing novel functions. *Methods Enzymol,* 497**,** 539-79.

HEJAZI, M. A., HOLWERDA, E. & WIJFFELS, R. H. 2004. Milking microalga Dunaliella salina for β-carotene production in two-phase bioreactors. *Biotechnology and Bioengineering,* 85**,** 475-481.

HELLIER, P., AL-HAJ, L., TALIBI, M., PURTON, S. & LADOMMATOS, N. 2013. Combustion and emissions characterization of terpenes with a view to their biological production in cyanobacteria. *Fuel*.

HELLINGWERF, K. J. & TEIXEIRA DE MATTOS, M. J. 2009. Alternative routes to biofuels: Light-driven biofuel formation from CO2 and water based on the [`]photanol' approach. *Journal of Biotechnology,* 142**,** 87-90.

HILLSON, N. J., ROSENGARTEN, R. D. & KEASLING, J. D. 2011. j5 DNA Assembly Design Automation Software. *ACS Synthetic Biology,* 1**,** 14-21.

HOICZYK, E. & HANSEL, A. 2000. Cyanobacterial Cell Walls: News from an Unusual Prokaryotic Envelope. *Journal of Bacteriology,* 182**,** 1191-1199.

HONG, H., PATEL, D. R., TAMM, L. K. & VAN DEN BERG, B. 2006. The outer membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel. *J Biol Chem,* 281**,** 7568-77.

HOWITT, C. A., UDALL, P. K. & VERMAAS, W. F. 1999. Type 2 NADH dehydrogenases in the cyanobacterium Synechocystis sp. strain PCC 6803 are involved in regulation rather than respiration. *J Bacteriol,* 181**,** 3994-4003.

HU, P., BORGLIN, S., KAMENNAYA, N. A., CHEN, L., PARK, H., MAHONEY, L., KIJAC, A., SHAN, G., CHAVARRÍA, K. L., ZHANG, C., QUINN, N. W. T., WEMMER, D., HOLMAN, H.-Y. & JANSSON, C. 2013. Metabolic phenotyping of the cyanobacterium Synechocystis 6803 engineered for production of alkanes and free fatty acids. *Applied Energy,* 102**,** 850-859.

HU, Q., SOMMERFELD, M., JARVIS, E., GHIRARDI, M., POSEWITZ, M., SEIBERT, M. & DARZINS, A. 2008. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *Plant Journal,* 54**,** 621-639.

HUANG, H.-H. & LINDBLAD, P. 2013. Wide-dynamic-range promoters engineered for cyanobacteria. *Journal of Biological Engineering,* 7**,** 10.

HUANG, H., CAMSUND, D., LINDBLAD, P. & HEIDORN, T. 2010. Design and characterization of molecular tools for a Synthetic Biology approach towards developing cyanobacterial biotechnology. *Nucleic Acids Res,* 38**,** 2577 - 2593.

HUSSEINY, M. I. & HENSEL, M. 2005. Rapid method for the construction of Salmonella enterica Serovar Typhimurium vaccine carrier strains. *Infect Immun,* 73**,** 1598 - 1605.

IIJIMA, Y., GANG, D. R., FRIDMAN, E., LEWINSOHN, E. & PICHERSKY, E. 2004. Characterization of geraniol synthase from the peltate glands of sweet basil. *Plant Physiology,* 134**,** 370-379.

IKEUCHI, M. & TABATA, S. 2001. Synechocystis sp PCC 6803 - a useful tool in the study of the genetics of cyanobacteria. *Photosynthesis Research,* 70**,** 73-83.

ILLMAN, A. M., SCRAGG, A. H. & SHALES, S. W. 2000. Increase in Chlorella strains calorific values when grown in low nitrogen medium. *Enzyme and Microbial Technology,* 27**,** 631-635.
INUI, M., SUDA, M., KIMURA, S., YASUDA, K., SUZUKI, H., TODA, H., YAMAMOTO, S., OKINO, S., SUZUKI, N. & YUKAWA, H. 2008. Expression of Clostridium acetobutylicum butanol synthetic genes in Escherichia coli. *Appl Microbiol Biotechnol,* 77**,** 1305-16.
ISKEN, S. & DE BONT, J. A. M. 1998. Bacteria tolerant to organic solvents. *Extremophiles,* 2**,** 229-238.
JANG, H. J., YOON, S. H., RYU, H. K., KIM, J. H., WANG, C. L., KIM, J. Y., OH, D. K. & KIM, S. W. 2011. Retinoid production using metabolically engineered Escherichia coli with a two-phase culture system. *Microb Cell Fact,* 10**,** 59.
JANSSON, C., DEBUS, R. J., OSIEWACZ, H. D., GUREVITZ, M. & MCINTOSH, L. 1987. Construction of an Obligate Photoheterotrophic Mutant of the Cyanobacterium Synechocystis 6803 : Inactivation of the psbA Gene Family. *Plant Physiology,* 85**,** 1021-1025.
JOJIMA, T., INUI, M. & YUKAWA, H. 2008. Production of isopropanol by metabolically engineered Escherichia coli. *Appl Microbiol Biotechnol,* 77**,** 1219-24.
JULSING, M. K., SCHREWE, M., CORNELISSEN, S., HERMANN, I., SCHMID, A. & BÜHLER, B. 2012. Outer Membrane Protein AlkL Boosts Biocatalytic Oxyfunctionalization of Hydrophobic Substrates in Escherichia coli. *Applied and Environmental Microbiology,* 78**,** 5724-5733.
KALSCHEUER, R., STÖLTING, T. & STEINBÜCHEL, A. 2006. Microdiesel: Escherichia coli engineered for fuel production. *Microbiology,* 152**,** 2529-2536.
KANEKO, T., SATO, S., KOTANI, H., TANAKA, A., ASAMIZU, E., NAKAMURA, Y., MIYAJIMA, N., HIROSAWA, M., SUGIURA, M. & SASAMOTO, S. 1996. Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res,* 3**,** 109 - 136.
KAY, R. A. & BARTON, L. L. 1991. Microalgae as food and supplement. *Critical Reviews in Food Science and Nutrition,* 30**,** 555-573.
KEASLING, J. D. 2008. Synthetic biology for synthetic chemistry. *ACS Chem Biol,* 3**,** 64-76.
KEASLING, J. D. & CHOU, H. 2008. Metabolic engineering delivers next-generation biofuels. *Nat Biotech,* 26**,** 298-299.
KNOOP, H., ZILLIGES, Y., LOCKAU, W. & STEUER, R. 2010. The metabolic network of Synechocystis sp. PCC 6803: systemic properties of autotrophic growth. *Plant Physiol,* 154**,** 410 - 422.
KOFFAS, M. A. G. 2009. Expanding the repertoire of biofuel alternatives through metabolic pathway evolution. *Proceedings of the National Academy of Sciences,* 106**,** 965-966.
KOK, M., OLDENHUIS, R., VAN DER LINDEN, M. P., MEULENBERG, C. H., KINGMA, J. & WITHOLT, B. 1989. The Pseudomonas oleovorans alkBAC

operon encodes two structurally related rubredoxins and an aldehyde dehydrogenase. *J Biol Chem,* 264**,** 5442-51.
KOKSHAROVA, O., C. WOLK 2002. Genetic tools for cyanobacteria. *Applied Microbiology and Biotechnology,* 58**,** 123.
KUFRYK, GALYNA, I., SACHET, MONIKA, SCHMETTERER, GEORG, VERMAAS & WIM, F. J. 2002. Transformation of the cyanobacterium *Synechocystis* sp. PCC 6803 as a tool for genetic mapping: optimization of efficiency. 206**,** 5.
KULKARNI, M. G. & DALAI, A. K. 2006. Waste Cooking OilAn Economical Source for Biodiesel: A Review. *Industrial & Engineering Chemistry Research,* 45**,** 2901-2913.
KUNERT, A., HAGEMANN, M. & ERDMANN, N. 2000. Construction of promoter probe vectors for Synechocystis sp. PCC 6803 using the light-emitting reporter systems Gfp and LuxAB. *J Microbiol Methods,* 41**,** 185-94.
LABARRE, J., CHAUVAT, F. & THURIAUX, P. 1989. Insertional mutagenesis by random cloning of antibiotic resistance genes into the genome of the cyanobacterium Synechocystis strain PCC 6803. *Journal of Bacteriology,* 171**,** 3449-3457.
LADYGINA, N., DEDYUKHINA, E. & VAINSHTEIN, M. 2006. A review on microbial synthesis of hydrocarbons. *Process Biochem,* 41**,** 1001 - 1014.
LAEMMLI, U. K. 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature,* 227**,** 680-685.
LAGARDE, D., BEUF, L. & VERMAAS, W. 2000. Increased Production of Zeaxanthin and Other Pigments by Application of Genetic Engineering Techniques toSynechocystis sp. Strain PCC 6803. *Applied and Environmental Microbiology,* 66**,** 64-72.
LAN, E. I. & LIAO, J. C. 2011. Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide. *Metab Eng,* 13**,** 353-63.
LANDRY, B. P., STÖCKEL, J. & PAKRASI, H. B. 2013. Use of Degradation Tags To Control Protein Levels in the Cyanobacterium Synechocystis sp. Strain PCC 6803. *Applied and Environmental Microbiology,* 79**,** 2833-2835.
LANGE, B. M., RUJAN, T., MARTIN, W. & CROTEAU, R. 2000. Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 13172-13177.
LEA-SMITH, D. J., ROSS, N., ZORI, M., BENDALL, D. S., DENNIS, J. S., SCOTT, S. A., SMITH, A. G. & HOWE, C. J. 2013. Thylakoid Terminal Oxidases Are Essential for the Cyanobacterium Synechocystis sp. PCC 6803 to Survive Rapidly Changing Light Intensities. *Plant Physiology,* 162**,** 484-495.
LEE, P. C. & SCHMIDT-DANNERT, C. 2002. Metabolic engineering towards biotechnological production of carotenoids in microorganisms. *Appl Microbiol Biotechnol,* 60**,** 1-11.
LEE, S. K., CHOU, H., HAM, T. S., LEE, T. S. & KEASLING, J. D. 2008. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current Opinion in Biotechnology,* 19**,** 556.

LEONARD, E., YAN, Y., FOWLER, Z. L., LI, Z., LIM, C.-G., LIM, K.-H. & KOFFAS, M. A. G. 2008. Strain Improvement of Recombinant Escherichia coli for Efficient Production of Plant Flavonoids. *Molecular Pharmaceutics,* 5**,** 257-265.

LESIC, B., BACH, S., GHIGO, J. M., DOBRINDT, U., HACKER, J. & CARNIEL, E. 2004. Excision of the high-pathogenicity island of Yersinia pseudotuberculosis requires the combined actions of its cognate integrase and Hef, a new recombination directionality factor. *Mol Microbiol,* 52**,** 1337 - 1348.

LESIC, B. & RAHME, L. 2008. Use of the lambda Red recombinase system to rapidly generate mutants in Pseudomonas aeruginosa. *BMC Molecular Biology,* 9**,** 20.

LI, M. Z. & ELLEDGE, S. J. 2007. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Meth,* 4**,** 251-256.

LI, Y., HORSMAN, M., WU, N., LAN, C. Q. & DUBOIS-CALERO, N. 2008. Biofuels from microalgae. *Biotechnol Prog,* 24**,** 815-20.

LIBERTON, M., HOWARD BERG, R., HEUSER, J., ROTH, R. & PAKRASI, H. B. 2006. Ultrastructure of the membrane systems in the unicellular cyanobacterium Synechocystis sp. strain PCC 6803. *Protoplasma,* 227**,** 129-38.

LICHTENTHALER, H. K. 2000. Non-mevalonate isoprenoid biosynthesis: enzymes, genes and inhibitors. *Biochem Soc Trans,* 28**,** 785-9.

LINDBERG, P., PARK, S. & MELIS, A. 2010. Engineering a platform for photosynthetic isoprene production in cyanobacteria, using Synechocystis as the model organism. *Metabolic Engineering,* 12**,** 70-79.

LINDBLAD, P., LINDBERG, P., OLIVEIRA, P., STENSJÖ, K. & HEIDORN, T. 2012. Design, Engineering, and Construction of Photosynthetic Microbial Cell Factories for Renewable Solar Fuel Production. *AMBIO: A Journal of the Human Environment,* 41**,** 163-168.

LIU, T., VORA, H. & KHOSLA, C. 2010. Quantitative analysis and engineering of fatty acid biosynthesis in E. coli. *Metab Eng,* 12**,** 378-86.

LIU, X. & CURTISS, R. 2009. Nickel-inducible lysis system in Synechocystis sp. PCC 6803. *Proceedings of the National Academy of Sciences,* 106**,** 21550-21554.

LIU, X. Y., SHENG, J. & CURTISS, R. 2011. Fatty acid production in genetically modified cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 6899-6904.

LIU, Z. Y., WANG, G. C. & ZHOU, B. C. 2008. Effect of iron on growth and lipid accumulation in Chlorella vulgaris. *Bioresour Technol,* 99**,** 4717-22.

LÓPEZ-MAURY, L., GARCÍA-DOMÍNGUEZ, M., FLORENCIO, F. J. & REYES, J. C. 2002. A two-component signal transduction system involved in nickel sensing in the cyanobacterium Synechocystis sp. PCC 6803. *Molecular Microbiology,* 43**,** 247-256.

LOREDANA, P., PÉTER, B. K., ZOLTAN, M., ANDREA, F. & IMRE, V. 2008. Construction of bioluminescent cyanobacterial reporter strains for detection of nickel, cobalt and zinc. *FEMS Microbiology Letters,* 289**,** 258-264.

LU, J., SHEAHAN, C. & FU, P. 2011. Metabolic engineering of algae for fourth generation biofuels production. *Energy & Environmental Science,* 4**,** 2451-2466.
LU, X. 2010. A perspective: Photosynthetic production of fatty acid-based biofuels in genetically engineered cyanobacteria. *Biotechnology Advances,* 28**,** 742-746.
LU, X., VORA, H. & KHOSLA, C. 2008. Overproduction of free fatty acids in E. coli: implications for biodiesel production. *Metab Eng,* 10**,** 333-9.
MACHADO, I. M. P. & ATSUMI, S. 2012. Cyanobacterial biofuel production. *Journal of Biotechnology,* 162**,** 50-56.
MARTIN, D. M., FALDT, J. & BOHLMANN, J. 2004. Functional characterization of nine Norway Spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. *Plant Physiol,* 135**,** 1908-27.
MARTIN, V. J. J., PITERA, D. J., WITHERS, S. T., NEWMAN, J. D. & KEASLING, J. D. 2003. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nat Biotech,* 21**,** 796-802.
MATSUNAGA, T., TAKEYAMA, H. & NAKAMURA, N. 1990. Characterization of cryptic plasmids from marine cyanobacteria and construction of a hybrid plasmid potentially capable of transformation of marine cyanobacterium, Synechococcus sp., and its transformation. *Appl Biochem Biotechnol,* 24-25**,** 151-60.
MAZEL, D., HOUMARD, J., CASTETS, A. M. & TANDEAU DE MARSAC, N. 1990. Highly repetitive DNA sequences in cyanobacterial genomes. *J Bacteriol,* 172**,** 2755-61.
MCFADDEN, G. I. 1999. Endosymbiosis and evolution of the plant cell. *Curr Opin Plant Biol,* 2**,** 513-9.
MEIER, R. L. 1955. Biological cycles in the transformation of solar energy into useful fuels. *In:* DANIELS, F. A. D., J.A., (ed.) *Solar Energy Research* Madison, WI: University of Wisconsin Press.
MELIS, A. 2009. Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency. *Plant Science,* 177**,** 272-280.
MELIS, A. 2013. Carbon partitioning in photosynthesis. *Curr Opin Chem Biol,* 17**,** 453-6.
MILLER, N. J., SAMPSON, J., CANDEIAS, L. P., BRAMLEY, P. M. & RICE-EVANS, C. A. 1996. Antioxidant activities of carotenes and xanthophylls. *FEBS Lett,* 384**,** 240-2.
MIN, H. & SHERMAN, L. A. 2010a. Genetic transformation and mutagenesis via single-stranded DNA in the unicellular, diazotrophic cyanobacteria of the genus Cyanothece. *Applied and environmental microbiology,* 76**,** 7641-7645.
MIN, H. & SHERMAN, L. A. 2010b. Genetic Transformation and Mutagenesis via ssDNA in the Unicellular, Diazotrophic Cyanobacteria of the Genus Cyanothece. *Appl. Environ. Microbiol.***,** AEM.01456-10.
MOHAMED, A., ERIKSSON, J., OSIEWACZ, H. D. & JANSSON, C. 1993. Differential expression of the psbA genes in the cyanobacterium Synechocystis 6803. *Mol Gen Genet,* 238**,** 161-8.

MOHAMED, A. & JANSSON, C. 1989. Influence of light on accumulation of photosynthesis-specific transcripts in the cyanobacterium Synechocystis 6803. *Plant Molecular Biology,* 13**,** 693-700.
MOLINA GRIMA, E., BELARBI, E. H., ACIÉN FERNÁNDEZ, F. G., ROBLES MEDINA, A. & CHISTI, Y. 2003. Recovery of microalgal biomass and metabolites: process options and economics. *Biotechnology Advances,* 20**,** 491-515.
MORGAN, T. R., SHAND, J. A., CLARKE, S. M. & EATON-RYE, J. J. 1998. Specific Requirements for Cytochrome c-550 and the Manganese-Stabilizing Protein in Photoautotrophic Strains of Synechocystis sp. PCC 6803 with Mutations in the Domain Gly-351 to Thr-436 of the Chlorophyll-Binding Protein CP47†. *Biochemistry,* 37**,** 14437-14449.
MUKHOPADHYAY, A., REDDING, A. M., RUTHERFORD, B. J. & KEASLING, J. D. 2008. Importance of systems biology in engineering microbes for biofuel production. *Current Opinion in Biotechnology,* 19**,** 228-234.
MULLINEAUX, C. W. 1999. The thylakoid membranes of cyanobacteria: structure, dynamics and function. *Functional Plant Biology,* 26**,** 671-677.
MURPHY, J. D. & POWER, N. M. 2009. An argument for using biomethane generated from grass as a biofuel in Ireland. *Biomass and Bioenergy,* 33**,** 504-512.
NARAYANAN, A., RIDILLA, M. & YERNOOL, D. A. 2011. Restrained expression, a method to overproduce toxic membrane proteins by exploiting operator-repressor interactions. *Protein Sci,* 20**,** 51-61.
NG, W., ZENTELLA, R., WANG, Y., TAYLOR, J. & PAKRASI, H. 2000. PhrA, the major photoreactivating factor in the cyanobacterium Synechocystis sp. strain PCC 6803 codes for a cyclobutane-pyrimidine-dimer-specific DNA photolyase. *Arch Microbiol,* 173**,** 412 - 417.
NGO, T. H. & SCHUMPE, A. 2012. Oxygen Absorption into Stirred Emulsions of n-Alkanes. *International Journal of Chemical Engineering,* 2012**,** 7.
NIELSEN, D. R., LEONARD, E., YOON, S. H., TSENG, H. C., YUAN, C. & PRATHER, K. L. 2009. Engineering alternative butanol production platforms in heterologous bacteria. *Metab Eng,* 11**,** 262-73.
NIKAIDO, H. 2003. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev,* 67**,** 593-656.
NOGALES, J., GUDMUNDSSON, S., KNIGHT, E. M., PALSSON, B. O. & THIELE, I. 2012. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proceedings of the National Academy of Sciences,* 109**,** 2678-2683.
OMATA, T. & MURATA, N. 1984. Isolation and characterization of three types of membranes from the cyanobacterium (blue-green alga) Synechocystis PCC 6714. *Archives of Microbiology,* 139**,** 113-116.
OSWALD, M., FISCHER, M., DIRNINGER, N. & KARST, F. 2007. Monoterpenoid biosynthesis in Saccharomyces cerevisiae. *Fems Yeast Research,* 7**,** 413-421.
OSWALD, W. J. & GOLUEKE, C. G. 1960. Biological transformation of solar energy. *Adv Appl Microbiol,* 2**,** 223-62.

PARMAR, A., SINGH, N. K., PANDEY, A., GNANSOUNOU, E. & MADAMWAR, D. 2011. Cyanobacteria and microalgae: A positive prospect for biofuels. *Bioresource Technology,* 102**,** 10163-10172.
PERALTA-YAHYA, P. P. & KEASLING, J. D. 2010. Advanced biofuel production in microbes. *Biotechnology Journal,* 5**,** 147-162.
PERALTA-YAHYA, P. P., ZHANG, F., DEL CARDAYRE, S. B. & KEASLING, J. D. 2012. Microbial engineering for the production of advanced biofuels. *Nature,* 488**,** 320-328.
PIMENTEL, D. & PATZEK, T. 2005. Ethanol Production Using Corn, Switchgrass, and Wood; Biodiesel Production Using Soybean and Sunflower. *Natural Resources Research,* 14**,** 65-76.
POLIQUIN, K., ERSHOV, Y. V., CUNNINGHAM, F. X., JR., WORETA, T. T., GANTT, R. R. & GANTT, E. 2004. Inactivation of sll1556 in Synechocystis strain PCC 6803 impairs isoprenoid biosynthesis from pentose phosphate cycle substrates in vitro. *J Bacteriol,* 186**,** 4685-93.
PORTER, R. D. 1986. Transformation in cyanobacteria. *Crc Critical Reviews in Microbiology,* 13**,** 111-132.
PORTER, R. D. 1988. [78] DNA transformation. *In:* LESTER PACKER, A. N. G. (ed.) *Methods in Enzymology.* Academic Press.
PRENTKI, P., BINDA, A. & EPSTEIN, A. 1991. Plasmid vectors for selecting IS1-promoted deletions in cloned DNA: sequence analysis of the omega interposon. *Gene,* 103**,** 17-23.
PROSPERI, C. 2000. Cyanobacteria in human affairs. *Intercienia,* 25**,** 303-306.
QUAN, J. & TIAN, J. 2011. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protocols,* 6**,** 242-251.
RADMER, R. & PARKER, B. 1994. Commercial applications of algae: opportunities and constraints. *Journal of Applied Phycology,* 6**,** 93-98.
RAGAUSKAS, A. J., WILLIAMS, C. K., DAVISON, B. H., BRITOVSEK, G., CAIRNEY, J., ECKERT, C. A., FREDERICK, W. J., HALLETT, J. P., LEAK, D. J., LIOTTA, C. L., MIELENZ, J. R., MURPHY, R., TEMPLER, R. & TSCHAPLINSKI, T. 2006. The Path Forward for Biofuels and Biomaterials. *Science,* 311**,** 484-489.
RAO, A. V. & AGARWAL, S. 1999. Role of lycopene as antioxidant carotenoid in the prevention of chronic diseases: A review. *Nutrition Research,* 19**,** 305-323.
RAVINDRAN, C. R., SUGUNA, S. & SHANMUGASUNDARAM, S. 2006. Electroporation as a tool to transfer the plasmid pRL489 in Oscillatoria MKU 277. *J Microbiol Methods,* 66**,** 174-6.
REINSVOLD, R. E., JINKERSON, R. E., RADAKOVITS, R., POSEWITZ, M. C. & BASU, C. 2011. The production of the sesquiterpene β-caryophyllene in a transgenic strain of the cyanobacterium Synechocystis. *Journal of Plant Physiology,* 168**,** 848-852.
REPPAS, N., RIDLEY, C., REPPAS, N., RIDLEY, C. & RODLEY, C. 2010. US: JOULE UNLIMITED INC.
RESCH, C. M. & GIBSON, J. 1983. Isolation of the carotenoid-containing cell wall of three unicellular cyanobacteria. *Journal of Bacteriology,* 155**,** 345-350.

RIPPKA R., A. H., M 1992. Catalouge of strains. *In:* PASTEUR, I. (ed.) *Pasture Culture Collection of Cyanobacterial Strains in Axenic Culture.* Paris.
RO, D.-K., PARADISE, E. M., OUELLET, M., FISHER, K. J., NEWMAN, K. L., NDUNGU, J. M., HO, K. A., EACHUS, R. A., HAM, T. S., KIRBY, J., CHANG, M. C. Y., WITHERS, S. T., SHIBA, Y., SARPONG, R. & KEASLING, J. D. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature,* 440**,** 940-943.
RO, D. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature,* 440**,** 940 - 3.
ROBERTSON, D., JACOBSON, S., MORGAN, F., BERRY, D., CHURCH, G. & AFEYAN, N. 2011. A new dawn for industrial photosynthesis. *Photosynthesis Research,* 107**,** 269-277.
ROSSI, M. S., PAQUELIN, A., GHIGO, J. M. & WANDERSMAN, C. 2003. Haemophore-mediated signal transduction across the bacterial cell envelope in Serratia marcescens: the inducer and the transported substrate are different molecules. *Mol Microbiol,* 48**,** 1467 - 1480.
ROUHIAINEN, L., PAULIN, L., SUOMALAINEN, S., HYYTIAINEN, H., BUIKEMA, W., HASELKORN, R. & SIVONEN, K. 2000. Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in Anabaena strain 90. *Mol Microbiol,* 37**,** 156-67.
RUDE, M. A. & SCHIRMER, A. 2009. New microbial fuels: a biotech perspective. *Current Opinion in Microbiology,* 12**,** 274.
RUFFING, A. M. 2011. Engineered cyanobacteria: Teaching an old bug new tricks. *Bioengineered,* 2**,** 136-149.
SACCHETTINI, J. C. & POULTER, C. D. 1997. Creating Isoprenoid Diversity. *Science,* 277**,** 1788-1789.
SAHA, R., VERSEPUT, A. T., BERLA, B. M., MUELLER, T. J., PAKRASI, H. B. & MARANAS, C. D. 2012. Reconstruction and comparison of the metabolic potential of cyanobacteria Cyanothece sp. ATCC 51142 and Synechocystis sp. PCC 6803. *PLoS One,* 7**,** e48285.
SAKAI, M., OGAWA, T., MATSUOKA, M. & FUKUDA, H. 1997. Photosynthetic conversion of carbon dioxide to ethylene by the recombinant cyanobacterium, Synechococcus sp. PCC 7942, which harbors a gene for the ethylene-forming enzyme of Pseudomonas syringae. *Journal of Fermentation and Bioengineering,* 84**,** 434-443.
SAMBROOK AND RUSSELL, D. W. 2001. *Molecular cloning,* New York, Cold Spring Harbor Laboratory Press.
SANDERSON, K. 2011. Lignocellulose: A chewy problem. *Nature,* 474**,** S12-S14.
SATO, K., KRIST, S. & BUCHBAUER, G. 2007. Antimicrobial effect of vapours of geraniol, (R)-(–)-linalool, terpineol, γ-terpinene and 1,8-cineole on airborne microbes using an airwasher. *Flavour and Fragrance Journal,* 22**,** 435-437.
SCHAKEL, S. F., BUZZARD, I. M. & GEBHARDT, S. E. 1997. Procedures for Estimating Nutrient Values for Food Composition Databases. *Journal of Food Composition and Analysis,* 10**,** 102-114.
SCHENK, P., THOMAS-HALL, S., STEPHENS, E., MARX, U., MUSSGNUG, J., POSTEN, C., KRUSE, O. & HANKAMER, B. 2008. Second Generation

Biofuels: High-Efficiency Microalgae for Biodiesel Production. *BioEnergy Research,* 1**,** 20-43.
SCHIRMER, A., RUDE, M. A., LI, X., POPOVA, E. & DEL CARDAYRE, S. B. 2010. Microbial Biosynthesis of Alkanes. *Science,* 329**,** 559-562.
SCHULZE, K., LOPEZ, D., TILLICH, U. & FROHME, M. 2011. A simple viability analysis for unicellular cyanobacteria using a new autofluorescence assay, automated microscopy, and ImageJ. *BMC Biotechnology,* 11**,** 118.
SCHWARTZ, W. 1975. G. E. Fogg, W. D. P. Stewart, P. Fay And A. E. Walsby, The Blue-Green Algae. VII, 459 S., 170 Abb. und 8 farbige Abb., 33 Tab., 1 Farbtafel. London-New York 1973: Academic Press, £ 8.50. *Zeitschrift für allgemeine Mikrobiologie,* 15**,** 68-68.
SCHWARZ, D., ORF, I., KOPKA, J. & HAGEMANN, M. 2013. Recent Applications of Metabolomics Toward Cyanobacteria. *Metabolites,* 3**,** 72-100.
SHEEHAN, J. 2009. Engineering direct conversion of CO2 to biofuel. *Nat Biotech,* 27**,** 1128-1129.
SHESTAKOV, S. V. & KHYEN, N. T. 1970. Evidence for genetic transformation in blue-green alga Anacystis nidulans. *Mol Gen Genet,* 107**,** 372-5.
SIELAFF, H., CHRISTIANSEN, G. & SCHWECKE, T. 2006. Natural products from cyanobacteria: Exploiting a new source for drug discovery. *IDrugs,* 9**,** 119-27.
SIKKEMA, J., DE BONT, J. A. & POOLMAN, B. 1995. Mechanisms of membrane toxicity of hydrocarbons. *Microbiological Reviews,* 59**,** 201-22.
SIMON, J. R. & MOORE, P. D. 1987. Homologous recombination between single-stranded DNA and chromosomal genes in Saccharomyces cerevisiae. *Molecular and Cellular Biology,* 7**,** 2329-2334.
SLEYTR, U. B., MESSNER, P., PUM, D. & SÁRA, M. 1993. Crystalline bacterial cell surface layers. *Molecular Microbiology,* 10**,** 911-916.
SOMOLINOS, M., GARCÍA, D., CONDÓN, S., MACKEY, B. M. & PAGÁN, R. 2010. Inactivation of Escherichia coli by citral. *Journal of Applied Microbiology,* 108**,** 1928-1939.
SONG, L. 2006. A soluble form of phosphatase in Saccharomyces cerevisiae capable of converting farnesyl diphosphate into E,E-farnesol. *Appl Biochem Biotechnol,* 128**,** 149-58.
SPOLAORE, P., JOANNIS-CASSAN, C., DURAN, E. & ISAMBERT, A. 2006. Commercial applications of microalgae. *J Biosci Bioeng,* 101**,** 87-96.
STANIER, R. Y., KUNISAWA, R., MANDEL, M. & COHEN-BAZIRE, G. 1971. Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriol Rev,* 35**,** 171-205.
STEEN, E., CHAN, R., PRASAD, N., MYERS, S., PETZOLD, C., REDDING, A., OUELLET, M. & KEASLING, J. 2008. Metabolic engineering of Saccharomyces cerevisiae for the production of n-butanol. *Microbial Cell Factories,* 7**,** 36.
STEEN, E. J., KANG, Y., BOKINSKY, G., HU, Z., SCHIRMER, A., MCCLURE, A., DEL CARDAYRE, S. B. & KEASLING, J. D. 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature,* 463**,** 559-562.

STEPHANOPOULOS, G. & SINSKEY, A. J. 1993. Metabolic engineering--methodologies and future prospects. *Trends Biotechnol,* 11**,** 392-6.
STUCKEN, K., ILHAN, J., ROETTGER, M., DAGAN, T. & MARTIN, W. F. 2012. Transformation and conjugal transfer of foreign genes into the filamentous multicellular cyanobacteria (subsection V) Fischerella and Chlorogloeopsis. *Curr Microbiol,* 65**,** 552-60.
SUMMONS, R. E., JAHNKE, L. L., HOPE, J. M. & LOGAN, G. A. 1999. 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature,* 400**,** 554-7.
SUN, W., WANG, S. & CURTISS, R. 2008. Highly Efficient Method for Introducing Successive Multiple Scarless Gene Deletions and Markerless Gene Insertions into the Yersinia pestis Chromosome. *Applied and Environmental Microbiology,* 74**,** 4241-4245.
TAN, L. T. 2007. Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochemistry,* 68**,** 954-979.
TAN, X., YAO, L., GAO, Q., WANG, W., QI, F. & LU, X. 2011. Photosynthesis driven conversion of carbon dioxide to fatty alcohols and hydrocarbons in cyanobacteria. *Metab Eng,* 13**,** 169 - 176.
TANDEAU DE MARSAC, N., BORRIAS, W. E., KUHLEMEIER, C. J., CASTETS, A. M., VAN ARKEL, G. A. & VAN DEN HONDEL, C. A. 1982. A new approach for molecular cloning in cyanobacteria: cloning of an Anacystis nidulans met gene using a Tn901-induced mutant. *Gene,* 20**,** 111-9.
TARONCHER-OLDENBURG, G. & STEPHANOPOULOS, G. 2000. Targeted, PCR-based gene disruption in cyanobacteria: inactivation of the polyhydroxyalkanoic acid synthase genes in Synechocystis sp. PCC6803. *Applied microbiology and biotechnology,* 54**,** 677-680.
TAYLOR, L. A. & ROSE, R. E. 1988. A correction in the nucleotide sequence of the Tn903 kanamycin resistance determinant in pUC4K. *Nucleic Acids Res,* 16**,** 358.
TEMSAMANI, J., KUBERT, M. & AGRAWAL, S. 1995. Sequence identity of the n-1 product of a synthetic oligonucleotide. *Nucleic Acids Research,* 23**,** 1841-1844.
THIEL, T. 2004. Genetic Analysis of Cyanobacteria. *In:* BRYANT, D. (ed.) *The Molecular Biology of Cyanobacteria.* Springer Netherlands.
THIEL, T. & POO, H. 1989. Transformation of a filamentous cyanobacterium by electroporation. *Journal of bacteriology,* 171**,** 5743-5746.
TILCHE, A. & GALATOLA, M. 2008. The potential of bio-methane as bio-fuel/bio-energy for reducing greenhouse gas emissions: a qualitative assessment for Europe in a life cycle perspective. *Water Sci Technol,* 57**,** 1683-92.
TIWARI, D. N. 1978. THE HETEROCYSTS OF THE BLUE-GREEN ALGA NOSTOCHOPSIS LOBATUS: EFFECTS OF CULTURAL CONDITIONS. *New Phytologist,* 81**,** 653-656.
TSONG, T. Y. 1991. Electroporation of cell membranes. *Biophysical Journal,* 60**,** 297-306.
UGHY, B. & AJLANI, G. 2004. Phycobilisome rod mutants in Synechocystis sp. strain PCC6803. *Microbiology,* 150**,** 4147-4156.
UKITA, T. & IKEDA, H. 1996. Role of the recJ gene product in UV-induced illegitimate recombination at the hotspot. *J Bacteriol,* 178**,** 2362-7.

VACHHANI, A. K., IYER, R. K. & TULI, R. 1993. A Mobilizable Shuttle Vector for the Cyanobacterium Plectonema Boryanum. *Journal of General Microbiology,* 139**,** 569-573.
VAN BEILEN, J. B., EGGINK, G., ENEQUIST, H., BOS, R. & WITHOLT, B. 1992. DNA sequence determination and functional characterization of the OCT-plasmid-encoded alkJKL genes of Pseudomonas oleovorans. *Mol Microbiol,* 6**,** 3121-36.
VAN BEILEN, J. B., PANKE, S., LUCCHINI, S., FRANCHINI, A. G., RÖTHLISBERGER, M. & WITHOLT, B. 2001. Analysis of Pseudomonas putida alkane-degradation gene clusters and flanking insertion sequences: evolution and regulation of the alk genes. *Microbiology,* 147**,** 1621-1630.
VARMAN, A. M., XIAO, Y., PAKRASI, H. B. & TANG, Y. J. 2013. Metabolic engineering of Synechocystis sp. strain PCC 6803 for isobutanol production. *Appl Environ Microbiol,* 79**,** 908-14.
VERMAAS, W. 1996. Molecular genetics of the cyanobacterium<i>Synechocystis</i> sp. PCC 6803: Principles and possible biotechnology applications. *Journal of Applied Phycology,* 8**,** 263-273.
VERMAAS, W. F. J. 1998. Gene modifications and mutation mapping to study the function of photosystem II. *Photosynthesis: Molecular Biology of Energy Capture.*
VERMAAS, W. F. J. 2007. Targeted Genetic Modification of Cyanobacteria: New Biotechnological Applications. *Handbook of Microalgal Culture.* Blackwell Publishing Ltd.
VON HEIJNE, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res,* 14**,** 4683-90.
VRANOVA, E., COMAN, D. & GRUISSEM, W. 2012. Structure and dynamics of the isoprenoid pathway network. *Mol Plant,* 5**,** 318-33.
WANG, B., WANG, J., ZHANG, W. & MELDRUM, D. R. 2012. Application of synthetic biology in cyanobacteria and algae. *Frontiers in Microbiology,* 3.
WANG, C., YOON, S. H., JANG, H. J., CHUNG, Y. R., KIM, J. Y., CHOI, E. S. & KIM, S. W. 2011. Metabolic engineering of Escherichia coli for α-farnesene production. *Metabolic Engineering,* 13**,** 648-655.
WANG, W., LIU, X. & LU, X. 2013. Engineering cyanobacteria to improve photosynthetic production of alka(e)nes. *Biotechnology for Biofuels,* 6**,** 69.
WERNER, R. & MERGENHAGEN, D. 1998. Mating Type Determination of Chlamydomonas reinhardtii by PCR. *Plant Molecular Biology Reporter,* 16**,** 295-299.
WHITTON, B. & POTTS, M. 2000. Introduction to cyanobacteria. *Ecology of cyanobacteria.* Dordrecht: Kluwer Academic Publishers.
WIJFFELS, R. H. & BARBOSA, M. J. 2010. An Outlook on Microalgal Biofuels. *Science,* 329**,** 796-799.
WIJFFELS, R. H., BARBOSA, M. J. & EPPINK, M. H. M. 2010. Microalgae for the production of bulk chemicals and biofuels. *Biofuels, Bioproducts and Biorefining,* 4**,** 287-295.
WILLIAMS, J. 1988. Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methods in Synechocystis-6803. *Methods Enzymol,* 167**,** 766 - 778.

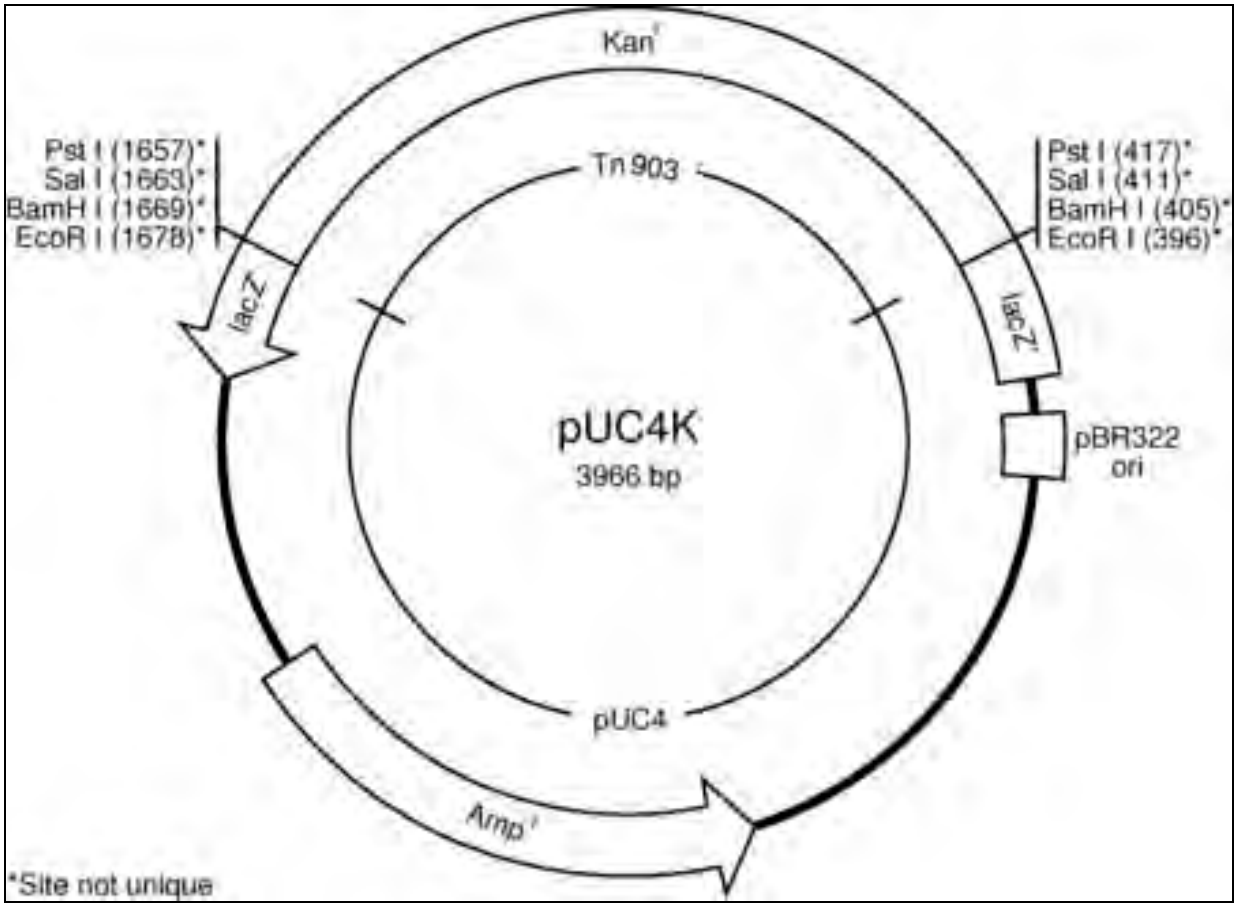
WILLIAMS, J. G. & SZALAY, A. A. 1983. Stable integration of foreign DNA into the chromosome of the cyanobacterium Synechococcus R2. *Gene,* 24**,** 37-51.
WITHERS, S. T., GOTTLIEB, S. S., LIEU, B., NEWMAN, J. D. & KEASLING, J. D. 2007. Identification of isopentenol biosynthetic genes from Bacillus subtilis by a screening method based on isoprenoid precursor toxicity. *Appl Environ Microbiol,* 73**,** 6277-83.
WOLK, C., ELHAI, J., KURITZ, T. & HOLLAND, D. 1993. Amplified expression of a transcriptional pattern formed during development of Anabaena. *Mol Microbiol,* 7**,** 441 - 445.
WOLK, C. P., VONSHAK, A., KEHOE, P. & ELHAI, J. 1984. Construction of shuttle vectors capable of conjugative transfer from Escherichia coli to nitrogen-fixing filamentous cyanobacteria. *Proceedings of the National Academy of Sciences,* 81**,** 1561-1565.
WU, Q. & VERMAAS, W. F. 1995. Light-dependent chlorophyll a biosynthesis upon chlL deletion in wild-type and photosystem I-less strains of the cyanobacterium Synechocystis sp. PCC 6803. *Plant Mol Biol,* 29**,** 933-45.
YANISCH-PERRON, C., VIEIRA, J. & MESSING, J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene,* 33**,** 103-19.
YOON, K. W., DOO, E. H., KIM, S. W. & PARK, J. B. 2008. In situ recovery of lycopene during biosynthesis with recombinant Escherichia coli. *J Biotechnol,* 135**,** 291-4.
YOON, S. H., PARK, H. M., KIM, J. E., LEE, S. H., CHOI, M. S., KIM, J. Y., OH, D. K., KEASLING, J. D. & KIM, S. W. 2007. Increased beta-carotene production in recombinant Escherichia coli harboring an engineered isoprenoid precursor pathway with mevalonate addition. *Biotechnol Prog,* 23**,** 599-605.
YOSHIDA, A., NISHIMURA, T., KAWAGUCHI, H., INUI, M. & YUKAWA, H. 2007. Efficient induction of formate hydrogen lyase of aerobically grown Escherichia coli in a three-step biohydrogen production process. *Appl Microbiol Biotechnol,* 74**,** 754-60.
YOSHIKUNI, Y., FERRIN, T. E. & KEASLING, J. D. 2006. Designed divergent evolution of enzyme function. *Nature,* 440**,** 1078-82.
YU, Y., YOU, L., LIU, D., HOLLINSHEAD, W., TANG, Y. & ZHANG, F. 2013. Development of Synechocystis sp. PCC 6803 as a Phototrophic Cell Factory. *Marine Drugs,* 11**,** 2894-2916.
ZANG, X., LIU, B., LIU, S., ARUNAKUMARA, K. K. & ZHANG, X. 2007. Optimum conditions for transformation of Synechocystis sp. PCC 6803. *J Microbiol,* 45**,** 241-5.
ZHANG, F., OUELLET, M., BATTH, T. S., ADAMS, P. D., PETZOLD, C. J., MUKHOPADHYAY, A. & KEASLING, J. D. 2012. Enhancing fatty acid production by the expression of the regulatory transcription factor FadR. *Metab Eng,* 14**,** 653-60.
ZHANG, K., SAWAYA, M. R., EISENBERG, D. S. & LIAO, J. C. 2008. Expanding metabolism for biosynthesis of nonnatural alcohols. *Proceedings of the National Academy of Sciences*.

ZHANG, Y.-H. P. & LYND, L. R. 2005. Cellulose utilization by Clostridium thermocellum: Bioenergetics and hydrolysis product assimilation. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 7321-7325.
ZHOU, J. & LI, Y. 2010. Engineering cyanobacteria for fuels and chemicals production. *Protein & Cell,* 1**,** 207-210.
ZHOU, J., ZHANG, H., ZHANG, Y., LI, Y. & MA, Y. 2012. Designing and creating a modularized synthetic pathway in cyanobacterium Synechocystis enables production of acetone from carbon dioxide. *Metab Eng,* 14**,** 394-400.
ZORIN, B., HEGEMANN, P. & SIZOVA, I. 2005. Nuclear-Gene Targeting by Using Single-Stranded DNA Avoids Illegitimate DNA Integration in Chlamydomonas reinhardtii. *Eukaryotic Cell,* 4**,** 1264-1272.

Appendices

Appendix 1: Plasmids utilised in this work.

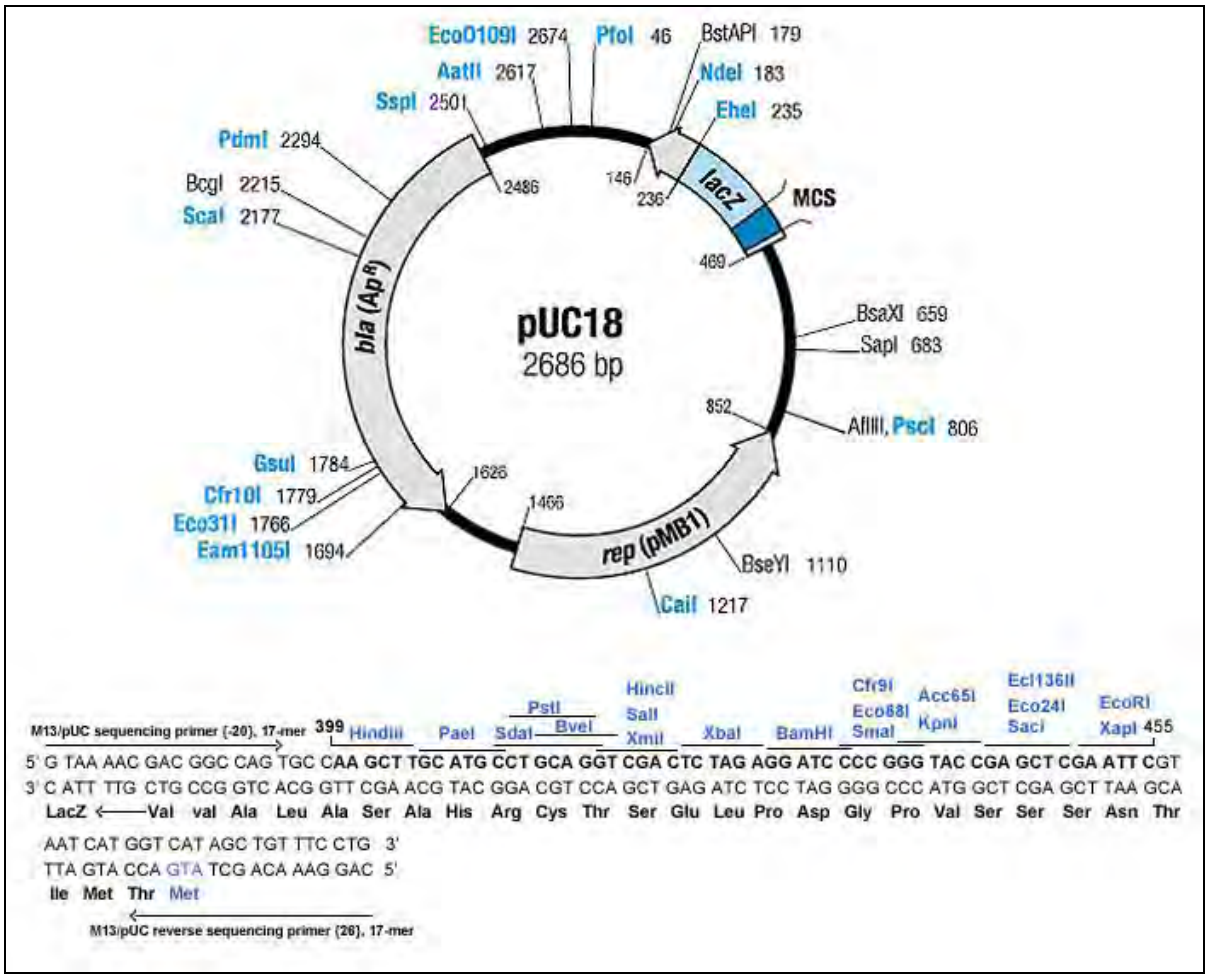
(1) pUK.4K: Used as a so.urce of the km^R cassette.



Taken from:

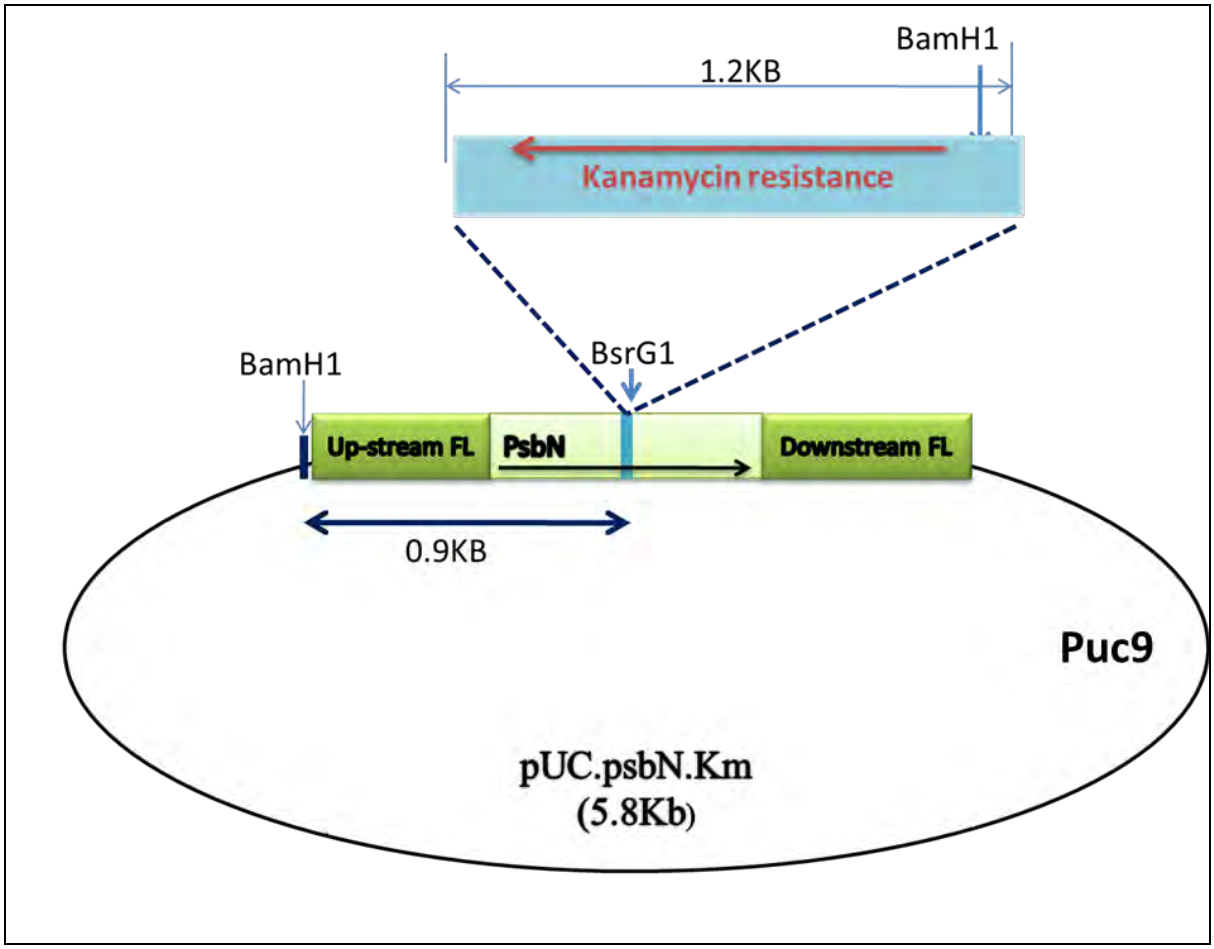
http://bric.postech.ac.kr/myboard/read.php?Board=exp_qna&id=26239

(2) pUC18: used as a base for the creation of the pLAH.A2 and pLAH.nrsB expression vectors.



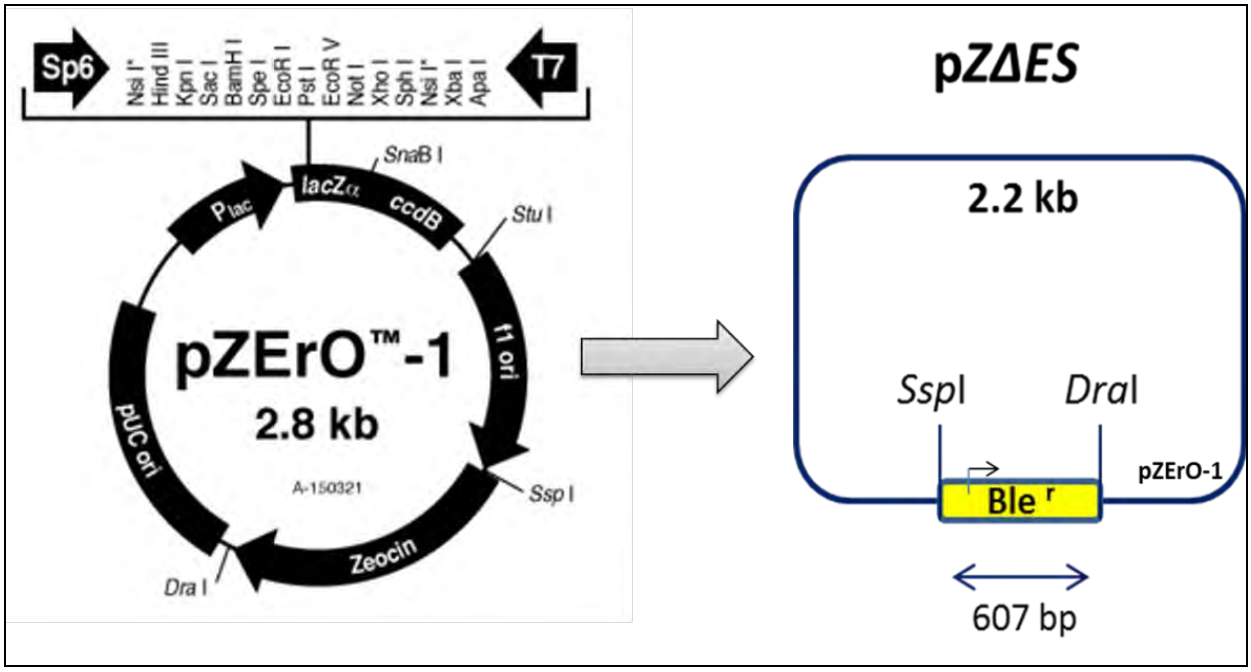
Taken from: http://www.genscript.com/vector/SD1162-pUC18_plasmid_DNA.html

(3) pUC.psbN.kmR (A): used for the generation of *psbN* knockouts.



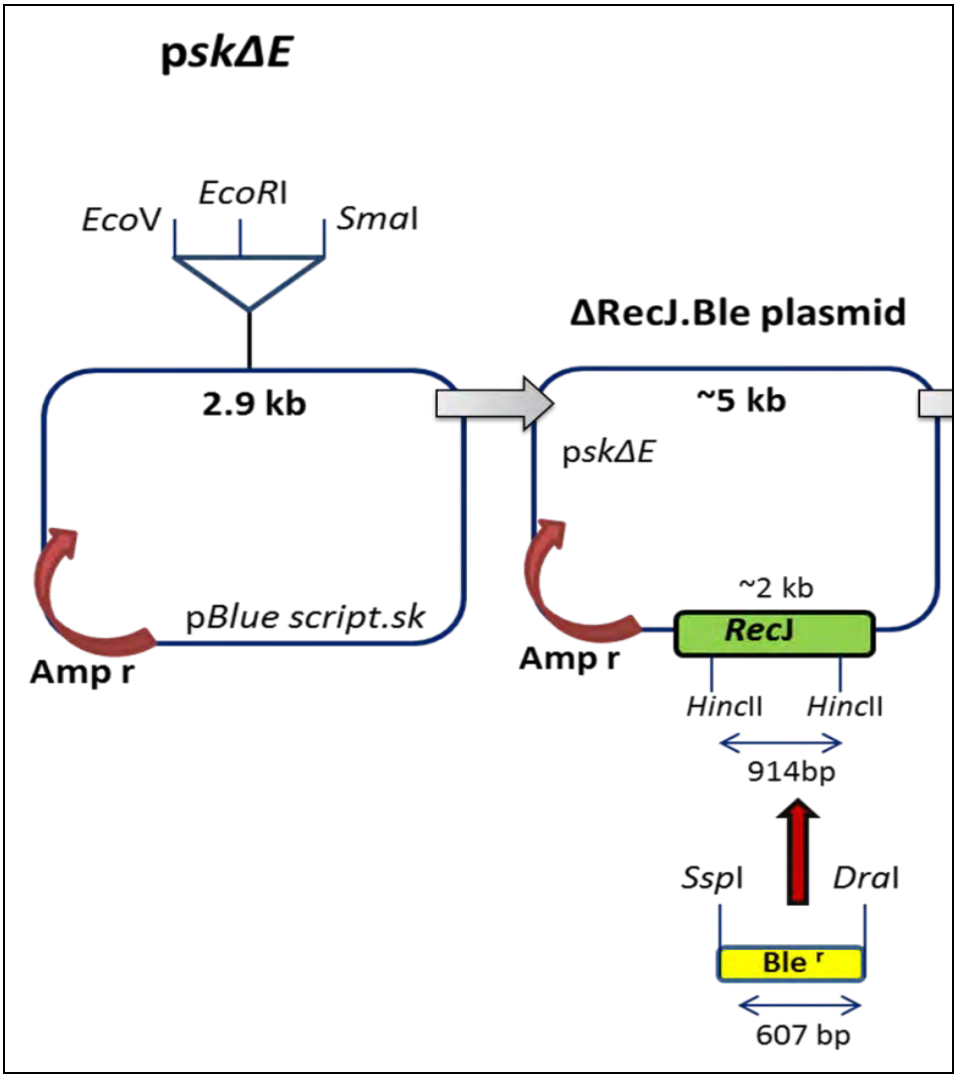
From: Saul Purton's personal collection.

(4) pZΔES: used as a source of the ble marker gene.



Note: A fragment of 541 bp (region between *Eco*RV and *Stu*I sites) was deleted from the p*ZErO-1* to remove toxic *ccdB* gene and create plasmid p*ZΔES*. The *Ble* fragment *(Ssp*I-*Dra*I fragment) = 607 bop. (Saul Purton's personal collection).

(5) pRecJ.ble: Used to create RecJ knockouts in *Synechocystis*.



From: Saul Purton's personal collection.

Appendix 2: Recipe for stock solutions preparation of BG11 medium

A) 100XBG11:	**g/L**
$NaNO_3$	149.6
$MgSO_4.7H_2O$	7.49
$CaCl_2$	3.60
Citric acid	0.60
Na_2EDTA 0.25M solution, pH 8.0	1.12 ml

B) Trace elements:	g/100ml
H_3BO_3	0.286
$MnCl_2.4H2O$	0.181
$ZnSO_4.7H2O$	0.022
$Na_2MoO_4.2H_2O$	0.039
$CuSO_4.5H_2O$	0.008
$Co(NO_3)2.6H_2O$	0.005

(c) Iron stock:	g/100ml
Ferric citrate	0.6
Or Ferric ammonium citrate	1.11

Stock solution	g/100ml
1) K_2HPO_4	3.05
2) Na2CO3	2.0
3) NaHCO3	8.401
4) TES buffer (NaOH to pH 8.2	22.9

Appendix 3: List of primers

Primer name	Primer sequence 5' to 3'	Function	Section
RecJ.F	5'-TACGCCATACCCTCATTTCC-3'	Testing for the isoaltion of + RecJ.Ble *Synechoystis* mutants	3.3.1
Recj.R	5'-CTAATTCCACCACCAGATCC-3'		
within km.F	5'-GGAAAGCCACGTTGTGTCTCAAAATCTC-3'	Amplify km cassette	3.3.3
within km.R	5'-GAAGAAGGTGTTGCTGACTCATACC-3'		
psbA2.speA.F	5'- CCTCTACACAGCCCAGAACTATGG-3'	Amplify area of interest for sequencing	3.3.3.4
psbA2.speA.R	5'-TAATGCTTTATCCGCCGGCACAGG-3'		
GDP-Mannose.F(1)	5'- GCACTTTGCCGAACATTTCC-3'	check integration in psbA2.speA area of genome	3.3.3.4
km.faceout.F	5'-ccttttaacagcgatcgcg-3'		
GDP-Mannose.F(2)	5'-GGTCAGTGTTAAAGGTAGAAG-3'	for sequanciny psbA2.speA area	3.3.3.4
km.faceout.new.F	5'-acaacgcagaccgttccgtgg-3'		
Ble from pZΔEs.F	5'- AATATTAACGTTTACAATTTCGCCTGATGC-3'	amplification ble cassette from the pZΔES plasmid	3.3.4
Ble from pZΔEs.R	5'-CACCTAGATCCTTTTAAATTAAAAATG-3'		
GDP-Mannose.F(1)	5'- GCACTTTGCCGAACATTTCC-3'	for amplification of the psbA2.speA area	3.3.6.2
5' Ble screen	5'-GATGGCCAAGCTGACCAGCGC-3'		
Primer name	**Primer sequence 5'to 3'**	**Function**	**Section**
PsbN.km.1000F	5'-GAAGCCAAAAATTCCGTCACC-3'	Amplify psbN.km knockout with 1000 flank homology on either side	4.3.1.1
PsbN.km.1000R	5'- TTTTACCCACACCTCCAATCC -3'		
PsbN.km.600F	5' - GGCCGGCTAAGTTACCGCTACC - 3'	Amplify psbN.km knockout with 600 flank homology on either side	4.3.1.1
PsbN.km.600R	5' - CCGGCTCCAATTAGGTAAGGCAACTGCC-3'		
PsbN.km.400F	5' - CCATTAACCAGCCCAGTCCACGC- 3'	Amplify psbN.km knockout with 400 flank homology on either side	4.3.1.1
PsbN.km.400R	5'- GTGTAACCAGTGGCTAAGTATGTGC- 3'		
PsbN.km.200F	5'- GAGTGCGCTGTGCCATAGACTTCC -3'	Amplify psbN.km knockout with 200 flank homology on either side	4.3.1.1
PsbN.km.200R	5'- GTTGGTGCAAGGAGATCCTTACTGGC - 3'		
PsbN.km.100F	5' - GACAACAGAAACAACTATCAATCTTACC-3'	Amplify psbN.km knockout with 100 flank homology on either side	4.3.1.1
PsbN.km.100R	5'-GTGTTTAAATTAAAGACCCTGACGG -3'		
PsbN.km.50F	5'-CGCAACAGTTCTTAGCATTACTTTTGC -3'	Amplify psbN.km knockout with 50 flank homology on either side	4.3.1.1
PsbN.km.50R	5'- CGTGGTCATCGAAGGGATCACCC - 3'		

PsbN.km.30F	5'-GCTGTCATTCTCATCGCCATCACTGG- 3'	Amplify psbN.km knockout with 30 flank homology on either side	4.3.1.1
PsbN.km.30R	5'-CTCAGCGGACGGAGGGCCAAAGG- 3'		
long tail.37.km	5'-gctgtcattctcatcgccatcactggtctggctgtgtGAAGAAGGTGTTGCTGACTCATACC-3'	Amplify psbN.km knockout with 30 flank homology using 1-step PCR	4.3.2.1
long tail.29.km	5'-ctcagcggacggagggccaaaggaagtgtGGAAAGCCACGTTGTGTCTCAAAATCTC-3'		
long tail.62.km	5'-cgcaacagttcttagcattacttttgctgtcattctcatcgccatcac tggtctggctgtgt GAAGAAGGTGTTGCTGACTCATACC-3'	Amplify psbN.km knockout with 50 flank homology using 1-step PCR	4.3.3.3
long tail.54.km	5'-cgtggtcatcgaagggatcacccaactcagcggacggagggccaaa ggaagtgtGGAAAGCCACGTTGTGTCTCAAAATCTC-3'		
psb28.2.check.F	5'-ggttggcaatgaactatttgtcc-3'	Check for correct integration in psb28.2 locus	4.3.3.3
psb28.2.check.R	5'-gcaaacaataataagggcaacagc-3'		
psbA2.check.F	5'-accacttgcttcatcattgcc-3'	Check for correct integration in psbA2 locus	4.3.3.3
psbA2.check.R	5'-aaccagatgccgattacagg-3'		
psbA2.spea.F	5'-cctctacacagcccagaactatgg-3'	check integration in psbA2.spea area of genome "hot spot"	4.3.3.3
km.faceout.R	5'-ccttttaacagcgatcgcg-3'		
LT62.ble.psbN.F	5'-cgcaacagttcttagcattacttttgctgtcattctcatcgccatcactggt ctggctgtgtGCATCTGTGCGGTATTTCACACCG-3'	Amplify psbN.ble knockout with ~50 flank homology using 1-step PCR	4.3.4.2
LT54.ble.psbN.R	5'-cgtggtcatcgaagggatcacccaactcagcggacggagggccaaa ggaagtgtTGAAGTTTTAGCACGTGTCAGTCC-3'		
LT52.ble.psbA2.F	5'-ttgctttacggtaacaacatcatctctggtgctgttgtaccttcttccaacg GCATCTGTGCGGTATTTCACACCG-3'	Amplify psbA2.ble knockout with ~50 flank homology using 1-step PCR	4.3.4.2
LT52.ble.psbA2.R	5'-ggaacggctgttgttgaaagaagcatattggaagatcaaccgaccaaa gtagTGAAGTTTTAGCACGTGTCAGTCC-3'		
LT53.ble.psb28.2F	5'- gttaaactgtcctgaccggctctagtaaaagttgttaggatcaaagaaatc ccGCATCTGTGCGGTATTTCACACCG-3'	Amplify psb28.2.ble knockout with ~50 flank homology using 1-step PCR	4.3.4.2
LT51.ble.psb28.2R	5'-ataatcaaggcttaatttcgctagattgtcccctgattagagtctgtttg cTGAAGTTTTAGCACGTGTCAGTCC-3'		
PsbN.km.1000F	5'- GAAGCCAAAAATTCCGTCACC -3'	Check for correct integration in psbN locus	4.3.4.4
3' Ble screen	5'-gtcctgctcctcggccacgaa-3' (internal Ble priner)		
psb28.2.check.F	5'-ggttggcaatgaactatttgtcc-3'	Check for correct integration in psb28.2 locus	4.3.4.4
3' Ble screen	5'-gtcctgctcctcggccacgaa-3' (internal Ble priner)		
psbA2.check.F	5'-accacttgcttcatcattgcc-3'	Check for correct integration in psbA2 locus	4.3.4.4
3' Ble screen	5'-gtcctgctcctcggccacgaa-3' (internal Ble priner)		

LT62.ble.psbN.F	5'-gctgtcattctcatcgccatcactggtctggctgtgtGCATCTGTGCGGTATTTCACACCG-3'	Amplify psbN.ble knockout with ~30 flank homology using 1-step PCR	4.3.4.4
LT54.ble.psbN.R	5'-ccaactcagcggacggaggccaaaggaagtgtTGAAGTTTTAGCACGTGTCAGTCC-3'		
LT52.ble.psbA2.F	5'-ctctggtgctgttgtaccttcttccaacgGCATCTGTGCGGTATTTCACACCG-3'	Amplify psbA2.ble knockout with ~30 flank homology using 1-step PCR	4.3.4.4
LT52.ble.psbA2.R	5'-gcatattggaagatcaaccgaccaaagtagTGAAGTTTTAGCACGTGTCAGTCC-3'		
LT53.ble.psb28.2F	5'-gctctagtaaaagttgttaggatcaaagaaatcccGCATCTGTGCGGTATTTCACACCG-3'	Amplify psb28.2.ble knockout with ~30 flank homology using 1-step PCR	4.3.4.4
LT51.ble.psb28.2R	5'-gctagattgtcccctgattagagtctgtttgcTGAAGTTTTAGCACGTGTCAGTCC-3'		
LT52.Phycobili.ble.F	5'ggtcaaaagagactcggcacccttgaggcgctcagaatccataatgtaatcg GCATCTGTGCGGTATTTCACACCG-3'	Amplify operon.ble knockout with ~50 flank homology using 1-step PCR	4.3.4.5
LT50.Phycobili.ble.R	5'cgtgacatggaaatcatcctccgctatgttacctacgcaaccttcaccggT GAAGTTTTAGCACGTGTCAGTCC -3'		
Phycobsome.check.F	5'-ttgccagacgactagaactgc-3'	Check for correct integration in the phycobilisome operon	4.3.4.5
Phycobsome.check.R	5'- ccgttgctgaaggcaacaaacgg-3'		
Primer name	**Primer sequence 5'to 3'**	**Function**	**Section**
psbA2-*Sph*I-1F	5´-ATGCATGCCCCTAAAAATATCAGAATCC-3´	Amplify upstream element in pLAH.A2 expression vector	5.3.1.1
psbA2-*Nde*I,*Sal*I-1R	5´-GAGAGTCGACGTCATATGGTTATAATTCC-3´		
psbA2-*Kpn*I-2F	5´-CGGTACCTTCCTTGGTGTAATGCC-3´	Amplify downstream element in pLAH.2 expression vector	5.3.1.1
psbA2-*EcoR*I-2R	5´-TCTAGAGAATTCTGGTCTTTGTAGCC-3´		
A2new-*Hind*III-F	5´-CAGTGTTAAAGCTTGAAGTTCGG-3´	Amplify upstream element in pLAH.nrsBR expression vector	5.3.1.2
A2new-*Hind*III-R	5´-CAGGAAGTTAATATACACGAAGC-3´		
nrsRegu.*Sph*I.F	5'-CCAGAAAGCCAATTGCATGCGACTACG-3'	Amplify *nrsB* promoter + regulatory element in pLAH.nrsB expression vector	5.3.1.2
nrsB.*Nde*I.*Xho*I.R	5'-CTACTCGAGTTGTTTTCATCATATGACCTCAAATTGGG-3'		
6803.nrsB.BleF	5'-GCCCATATGGCCAAGTTGACC-'3	Amplify 381 bp ble selectable marker with engineered *Nde*I and *BamH*I sites	5.3.1.3
6803.nrBb.Ble.R	5'-TGAGGATCCTCAGTCCTGCTCCTCGGCC-'3		
Ben.F.seq.ups	5'-CCCAGAACTATGGTAAAGGCG-3'	used to check correct integration of G.O.I in *psbA*2 site	5.3.1.3 + 5.3.4.3
out.GFP.AII.R	5'-GGTCTGCCTCGTGAAGAAGGTGTT-3'		
Farnesene.*Nde*I.F	5'-GCCCATATGGAATTTCGGGTGCATTTGC-3'	Amplify an HA tagged α–farnesene synthase gene	5.3.4.2
farnesene.HA.*BamH*I.R	5'-GTAGGATCCTTAGGCATAATCGGGCACATCAT AGGGATAGTTCACCAGGGGTTGAAAC-3'		
GES-*Nde*I-F	5'-GCCCATATGCGCCCACGCTTCTCTG-3'	Amplify GES gene from the pCRT7/CT-TOPO plasmid	5.3.2.3.i
GES-HA.tag-*BamH*I-R	5'-GTAGGATCCTTAGGCATAATCGGGCACA TCATAGGGATATTGAGTGAAGAAGAGGGCATC-3'		

internal.GES.R	5'-CCTGCAACCCCCAAATTTGACG-3'	check correct integration of GES in *psbA*2 site along with Ben.F.seq.usps primer	5.3.2.3.iii
Primer name	**Primer sequence 5'to 3'**	**Function**	**Section**
ALK.NdeI-F	5'-CTAGTCATAGGAGGTAAAACATATGAG-3'	Amplify alkL gene with its HA tag	6.3.1.1
ALK-Hatag-BamHI.R	5'- GTAGGATCCTTAGGCATAATCGGGCACATC ATAGGGATAGAAAACATACGACGCACCAAGAC-3'		
Ben.F.seq.ups	5'-CCCAGAACTATGGTAAAGGCG-3'	used to check correct integration of G.O.I in *psbA*2 site	6.3.1.2
out.GFP.AII.R	5'-GGTCTGCCTCGTGAAGAAGGTGTT-3'		

Appendix 4: CDS of GES

ACCESSION AY362553, Ocimum basilicum: geraniol synthase (GES) 1704 bp

ATGTCTTGTG CACGGATCAC CGTAACATTG CCGTATCGCT CCGCAAAAAC ATCAATTCAA CGGGGAATTA
CGCATTACCC CGCCCTTATA CGCCCACGCT TCTCTGCTTG CACGCCTTTG GCATCGGCGA TGCCTCTAAG
TTCAACTCCT CTCATCAACG GGGATAACTC TCAGCGTAAA AACACACGTC AACACATGGA GGAGAGCAGC
AGCAAGAGGA GAGAATATCT GCTGGAGGAA ACGACGCGAA AACTGCAGAG AAACGACACC GAATCGGTGG
AGAAACTCAA GCTTATCGAC AACATCCAAC AGTTGGGAAT CGGCTACTAT TTTGAGGACG CCATCAACGC
CGTACTCCGC TCGCCTTTCT CCACCGGAGA AGAAGACCTC TTCACCGCTG CTCTGCGCTT CCGCTTGCTC
CGCCACAACG GCATCGAAAT CAGCCCTGAA ATATTCCTAA AATTCAAGGA CGAGAGGGGA AAATTCGACG
AATCGGACAC GCTAGGGTTA CTGAGCTTGT ACGAAGCGTC AAATTTGGGG GTTGCAGGAG AAGAAATATT
GGAGGAGGCT ATGGAGTTTG CGGAGGCTCG CCTGAGACGG TCGCTGTCAG AGCCGGCGGC GCCGCTTCAT
GGTGAGGTGG CGCAAGCGCT AGATGTGCCG AGGCATCTGA GAATGGCGAG GTTGGAAGCG AGACGATTCA
TCGAGCAGTA TGGTAAACAG AGCGATCATG ATGGAGATCT TTTGGAGCTG GCAATTTTGG ATTATAATCA
AGTTCAGGCT CAACACCAAT CCGAACTCAC TGAAATAATC AGGTGGTGGA AGGAGCTCGG TTTGGTGGAT
AAGTTGAGTT TTGGGCGAGA CAGACCATTG GAGTGCTTTT TGTGGACCGT GGGGCTCCTC CCAGAGCCCA
AGTATTCGAG CGTTAGAATA GAGTTGGCGA AAGCCATCTC TATTCTCTTA GTGATCGATG ATATTTTCGA
TACCTATGGA GAGATGGATG ACCTCATCCT CTTCACCGAT GCAATTCGAA GATGGGATCT TGAAGCAATG
GAGGGGCTCC CTGAGTACAT GAAAATATGC TACATGGCGT TGTACAATAC CACCAATGAA GTATGCTACA
AAGTGCTCAG GGATACTGGA CGGATTGTCC TCCTTAACCT CAAATCTACG TGGATAGACA TGATTGAAGG
TTTCATGGAG GAAGCAAAAT GGTTCAATGG TGGAAGTGCA CCAAAATTGG AAGAGTATAT AGAGAATGGA
GTGTCCACGG CAGGAGCATA CATGGCTTTT GCACACATCT TCTTTCTCAT AGGAGAAGGT GTTACACACC
AAAATTCCCA ACTCTTCACC CAAAAACCCT ACCCCAAGGT CTTCTCCGCC GCCGGCCGCA TTCTTCGCCT
CTGGGATGAT CTCGGAACCG CCAAGGAAGA GCAAGAGCGA GGAGATCTGG CTTCGTGCGT GCAGTTATTT
ATGAAAGAGA AGTCGTTGAC GGAAGAGGAG GCAAGAAGTC GCATTTTGGA AGAGATAAAA GGATTATGGA
GGGATCTGAA TGGGGAACTG GTCTACAACA AGAATTTGCC GTTATCCATA ATCAAAGTCG CACTTAACAT
GGCGAGAGCT TCTCAAGTTG TGTACAAGCA CGATCAAGAC ACTTATTTTT CAAGCGTAGA CAATTATGTG
GATGCCCTCT TCTTCACTCA ATAA

Blue = predicted transit sequence

Appendix 5: Full sequence of RT17 (CT-TOPO) plasmid

>pCRT7/CT-TOPO, 2702 bp
GGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAAC
TTTAAGAAGGAATTGCCCTTAAGGGCAATTCGAAGCTTGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCT
ACGCGTACCGGTCATCATCACCATCACCATTGAGTTTAAACTATATAGAATAAAAGAAGAAACCTTAGCTGAGCAATAAC
TAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTAACGCTT
ACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTAT
CCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATGTGAGGAGGGCCACCATGGCCAAGTTGACCAGTGCCGTTC
CGGTGCTCACCGCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAG
GACGACTTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACAC
CCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCCGGGACG
CCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGC
GTGCACTTCGTGGCCGAGGAGCAGGACTGACACATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTT
ATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTT
TTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGT
CGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGT
AAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGA
AGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCC
ATACCAAACGACGAGAGTGACACCACGATGCCTGTAGCAATGCCAACAACGTTGCGCAAACTATTAACTGGCGAACTACT
TACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTC
CGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGAT
GGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGA
GATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTC
ATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTC
CACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCA
AACAAAAAAACGCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTC
AGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCC
TACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAA
GACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACC
TACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCC
GGTAAGCGGCAGGGTCGGAACAGGAGACGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACG
CGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGAT
AACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGA
AGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCA

Appendix 6: Annotated sequence of the RecJ.km.HincII transformant vrs WT6803 sequence.

RecJ.km.HincII transformant

Key: CDS of Sll1212 (GDP-mannose 4, 6 dehydratase) (green) - km^r cassette (grey) - KpnI site (orange) – *speA* (blue), Blue highlighted sequences are the HIPI palindromic sequences.

gaatggtaatttttggGCGATCGCcaccggctgcaattaattgggcaacaacttgtttaccgaggaaaccggccccgcca
gtgaccaaaatacgttgattttctagggaaagcatggttggttaagaaaatattgagttaagggagggatttgcctaggt
ttactgtgcccataaaattctccaccatttgaaggccagggaaaattgctagttgttaaacactttgatcgaggcaacag
ttaagacgctcaaccaagcttaatcgacggcatggccattttggctccgaatgaaggccatatcctgggccatgagctcc
ttgattcgtccactctggttgggagaagttaaaccaaggactgctaagtctgcctccaccattaaatgaactaattccgt
aaaggtaaccgagggttcccagcccaattgggctttggttttcgccggatcgccgattaataaatctacttcagcggggc
gtaaatagcgctcatcaaaggccacatagttctgccagttgaggttgacgtaaccaaaggcaatttctaaaaattccttc
acttcgtgggtttcccccgtggccaccacatagtcatcgggctgttcctgttgcaacatggcccacatggcccgtacata
gtccttggcatagccccaatcccgcttggaatcgatattgcctaaatacaatttcttttgggtgccggccacaattctgg
cgatcgcctagtaattttcctggttacaaaggtttctccccggcgggggggattcgtggttgaacaaaatgccgttacag
gcgaataagtcataggattcccgatagttcaccgtttgccaatggccataaaccttggcacaggcgtagggactgcgggg
ataaaaggggtggtttccttttggggaatctcctgcactttgccgaacatttccgaagaaccggcttgatagaacctta
cttggatgccggtgcgatgttgataatcccgaatcgcttccaatagtcgtagcgtccccatggccactgaatctacagtg
tattccggagaatcaaagctcacccgcacgtgggattgggcccccagattgtaaatctccgtcggtttgacatcttctaa
aatgcggcgcagggtggtgccgtcggtcagatcaccataatgaagtcggagtttcgcctcaagatcatggggatcaacat
aaagatgatcaatgcggtcagtgttaaaggtagaagttcggcgaatgatgccatggacttggtagccctttccaacaac
aattcactcagataggagccatcttgccccgtgatgcctgtcagcaaaacaactttagactttgacattagttaattttt
ccccattgccccaaaatacatccccctaaaaatatcagaatccttgcccagatgcaggccttctggcgatcgccatggtg
agcaacgattgcggctttagcgttccagtggatatttgctgggggttaatgaaacattgtggcggaacccagggacaatg
tgaccaaaaaattcagggatatcaataagtattaggtatatggatcataattgtatgcccgactattgcttaaactgact
gaccactgaccttaagagtaatggcgtgcaaggcccagtgatcaatttcattatttttcattatttcatctccattgtcc
ctgaaaatcagttgtgtcgcccctctacacagcccagaactatggtaaaggcgcacgaaaaaccgccaggtaaactcttc
tcaacccccaaaacgccctctgtttacccatggaaaaaacgacaattacaagaaagtaaaacttatgtcatctataagct
tGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGACctgcaggggggggggggcgctgaggtctgcctcgtgaagaaggt
gttgctgactcataccaggcctgaatcgccccatcatccagccagaaagtgagggagccacggttgatgagagctttgtt
gtaggtggaccagttggtgattttgaacttttgctttgccacggaacggtctgcgttgtcgggaagatgcgtgatctgat
ccttcaactcagcaaaagttcgatttattcaacaaagccgccgtcccgtcaagtcagcgtaatgctCtgccagtgttaca
accaattaaccaattctgaTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATA
CCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTA
TCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGA
AATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAG
CCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATAC
GCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATAT
TTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCA
TCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGT

AACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTG
TCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGC
CTCGAGCAAGACGTTTCCCGTTGAATATGGCTCATaacacccttgtattactgtttatgtaagcagacagttttattgt
tcatgatgatatatttttatcttgtgCAATGTAacatcagagattttgagACACAAcgtggctttcccccccccccctgc
agGTCgggtacctttccttggtgtaatgccaactgaataatctgcaaattgcactctccttcaatgggggtgctttttgc
ttgactgagtaatcttctgattgctgatcttgattgccatcgatcgccggggagtccggggcagttaccattagagagtc
tagagaattaatccatcttcgatagaggaattatggggaagaacctgtgccggcggataaagcattaggcaagaaattc
aagaaaaaaaatgcctcctggagcattgaagaaagcgaagctctgtaccgggttgaggcctgggggcaccttattttgc
cattaatgccgctggtaacataaccgtctctcccaacggcgatcggggcggttcgttagatttgttggaactggtggaag
ccctgcggcaaagaaagctcggcttacccctattaattcgtttttccgatattttggccgatcgcctagagcgattgaat
agttgttttgccaaggcgatcgccgttacaattaccccaacacctatcaggcggtttatccggtcaaatgtaaccagca
acgacatctggtggaagccctggttcgctttgggcaaacttcccagtgtggattggaggcaggttccaaaccggaattga
tgattgccctcgcaactctaccacctcccttagaccgtcaggacaagcataccaagcccctaatcatttgtaatggctac
aaagaccaggattatctagaaacagctctgttagccaaacgcttaggccatcgtcccatcatcatcattgaacaactacg

WT sequence

Key: CDS of Sll1212 (GDP-mannose 4, 6 dehydratase) (green) - CDS of psbAII (purple) with promoter (maroon and underlined) – *speA* (blue), Blue highlighted sequences are the HIPI palindromic sequences.

gaatggtaatttttggGCGATCGCcaccggctgcaattaattgggcaacaacttgtttaccgaggaaaccggccccgcca
gtgaccaaaatacgttgattttctagggaaagcatggttggttaagaaaatattgagttaagggagggatttgcctaggt
ttactgtgcccataaaattctccaccatttgaaggccagggaaaattgctagttgttaaacactttgatcgaggcaacag
ttaagacgctcaaccaagcttaatcgacggcatggccattttggctccgaatgaaggccatatcctgggccatgagctcc
ttgattcgtccactctggttgggagaagttaaaccaaggactgctaagtctgcctccaccattaaatgaactaattccgt
aaaggtaaccgagggttcccagcccaattgggctttggttttcgccggatcgccgattaataaatctacttcagcggggc
gtaaatagcgctcatcaaaggccacatagttctgccagttgaggttgacgtaaccaaaggcaatttctaaaaattccttc
acttcgtgggtttcccccgtggccaccacatagtcatcgggctgttcctgttgcaacatggcccacatggcccgtacata
gtccttggcatagccccaatcccgcttggaatcgatattgcctaaatacaatttcttttgggtgccggccacaattctgg
cgatcgcctagtaattttcctggttacaaaggtttctccccggcgggggattcgtggttgaacaaaatgccgttacag
gcgaataagtcataggattcccgatagttcaccgtttgccaatggccataaaccttggcacaggcgtagggactgcgggg
ataaaaggggtggtttccttttggggaatctcctgcactttgccgaacatttccgaagaaccggcttgatagaacctta
cttggatgccggtgcgatgttgataatcccgaatcgcttccaatagtcgtagcgtccccatggccactgaatctacagtg
tattccggagaatcaaagctcacccgcacgtgggattgggcccccagattgtaaatctccgtcggtttgacatcttctaa
aatgcggcgcagggtggtgccgtcggtcagatcaccataatgaagtcggagtttcgcctcaagatcatggggatcaacat
aaagatgatcaatgcggtcagtgttaaaggtagaagttcggcgaatgatgccatggacttggtagccctttttccaacaac
aattcactcagataggagccatcttgccccgtgatgcctgtcagcaaaacaactttagactttgacattagttaatttt
ccccattgccccaaaatacatccccctaaaaatatcagaatccttgcccagatgcaggccttctggcgatcgccatggtg
agcaacgattgcggctttagcgttccagtggatatttgctgggggttaatgaaacattgtggcggaacccagggacaatg
tgaccaaaaaattcagggatatcaataagtattaggtatatggatcataattgtatgcccgactattgcttaaactgact
gaccactgaccttaagagtaatggcgtgcaaggcccagtgatcaatttcattatttttcattatttcatctccattgtcc
ctgaaaatcagttgtgtcgcccctctacacagcccagaactatggtaaaggcgcacgaaaaaccgccaggtaaactcttc
tcaacccccaaaacgccctctgtttacccatggaaaaaacgacaattacaagaaagtaaaacttatgtcatctataagct
tcgtgtatattaacttcctgttacaaagctttacaaaactctcattaatcctttagactaagtttagtcagttccaatct
gaacatcgacaaatacataaggaattataaccaaatgacaacgactctccaacagcgcgaaagcgcttccttgtgggaac

agttttgtcagtgggtgacctctaccaacaaccggatttatgtcggttggttcggtaccttgatgatccccaccctctta
actgccaccacttgcttcatcattgccttcatcgccgctccccccgttgacatcgacggtatccgtgagcccgttgctgg
ttctttgctttacggtaacaacatcatctctggtgctgttgtaccttcttccaacgctatcggtttgcacttctacccca
tctgggaagccgcttccttagatgagtggttgtacaacggtggtccttaccagttggtagtattccacttcctcatcggc
attttctgctacatgggtcgtcagtgggaactttcctaccgcttaggtatgcgtccttggatttgtgtggcttactctgc
ccccgtatccgctgccaccgccgtattcttgatctaccccattggtcaaggctccttctctgatggtatgcccttgggta
tttctggtaccttcaacttcatgatcgtgttccaagctgagcacaacatcctgatgcacccctccacatgttaggtgtg
gctggtgtattcggtggtagcttgttctccgccatgcacggttccttggtaacctcctccttggtgcgtgaaaccaccga
agttgaatcccagaactacggttacaaattcggtcaagaagaagaaacctacaacatcgttgccgcccacggctactttg
gtcggttgatcttccaatatgcttctttcaacaacagccgttccttgcacttcttcttgggtgcttggcctgtaatcggc
atctggttcactgctatgggtgtaagcaccatggcgttcaacctgaacggtttcaacttcaaccagtccatcttggatag
ccaaggccgggtaatcggcacctgggctgatgtattgaaccgagccaacatcggttttgaagtaatgcacgaacgcaatg
cccacaacttccccctcgacttagcgtctggggagcaagctcctgtggctttgaccgctcctgctgtcaacggttaattc
cttggtgtaatgccaactgaataatctgcaaattgcactctccttcaatgggggtgcttttgcttgactgagtaatct
tctgattgctgatcttgattgccatcgatcgccggggagtccggggcagttaccattagagagtctagagaattaatcca
tcttcgatagaggaattatggggaagaacctgtgccggcggataaagcattaggcaagaaattcaagaaaaaaaatgcc
tcctggagcattgaagaaagcgaagctctgtaccgggttgaggcctgggggcaccttattttgccattaatgccgctgg
taacataaccgtctctcccaacggcgatcggggcggttcgttagatttgttggaactggtggaagccctgcggcaaagaa
agctcggcttacccctattaattcgtttttccgatattttggccgatcgcctagagcgattgaatagttgttttgccaag
gcgatcgccgttacaattaccccaacacctatcaggcggtttatccggtcaaatgtaaccagcaacgacatctggtgga
agccctggttcgctttgggcaaacttcccagtgtggattggaggcaggttccaaaccggaattgatgattgccctcgcaa
ctctaccacctcccttagaccgtcaggacaagcataccaagcccctaatcatttgtaatggctacaaagaccaggattat
ctagaaacagctctgttagccaaacgcttaggccatcgtcccatcatcatcattgaacaactacg

Appendix 7: Sequence results of LT50.psbA2.km^{R} PCR product

Alignment results between LT50.psbA2.km (Forward reaction sequence analysis) showing nucleotide deletions and mutations at the flanks homologous to the psbA2 gene.

```
LA_LT50_psbAI    190 GGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCT
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           801 GGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCT

LA_LT50_psbAI    240 TGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAAC
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           851 TGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAAC

LA_LT50_psbAI    290 GGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAG
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           901 GGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAG

LA_LT50_psbAI    340 TTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTT
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           951 TTTCATTTGATGCTCGATGAGTTTTTCTAAtcagaattggttaattggtt

LA_LT50_psbAI    390 GTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGGCTTTGTTG
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
Visual          1001 gtaacactggcaGagcattacgctgacttgacgggacggcggctttgttg

LA_LT50_psbAI    440 AATAAATCGAACTTTTGCTGAGTTGAAGGATCAGATCACGCATCTTCCCG
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
Visual          1051 aataaatcgaacttttgctgagttgaaggatcagatcacgcatcttcccg

LA_LT50_psbAI    490 ACAACGCAGACCGTTCCGTGGCAAAGCAAAAGTTCAAAATCACCAACTGG
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
Visual          1101 acaacgcagaccgttccgtggcaaagcaaaagttcaaaatcaccaactgg

LA_LT50_psbAI    540 TCCACCTACAACAAAGCTCTCATCAACCGTGGCTCCCTCACTTTCTGGCT
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
Visual          1151 tccacctacaacaaagctctcatcaaccgtggctccctcactttctggct

LA_LT50_psbAI    590 GGATGATGGGGCGATTCAGGCCTGGTATGAGTCAGCAACACCTTCTTCCT
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
Visual          1201 ggatgatggggcgattcaggcctggtatgagtcagcaacaccttcttcct

LA_LT50_psbAI    640 ACTTTGGTCGGTTGATCTTCCAATATGCTTCTTTC----ACATGGCCCTT
                     |||||||||||||||||||||||||||||||||||    |||  |||.||
Visual          1251 actttggtcggttgatcttccaatatgcttctttcaacaaca--gccgtt

LA_LT50_psbAI    686 TCCAAACCAACACGGGAGTCGAATGCAACCGGCGCAGGAACACTGCCAGC
                     .|
Visual          1299 cc-----------------------------------------------
```

Appendix 8: Operon responsible for phycobilisome expression

Key: sll1580 (orange) - sll1579 (green) - sll1578 (blue) - sll1577 (red), primers used for PCR check (purple). Primers used for one-step PCR amplification (underlined).

>Chr:724093..729092
cctagttttgttcgttgctgtcagcttgatagggacgaatgttaacaatagttccccccagacgggaaatgc
ggcgcatctcatcattcatacgggcatagggaactttgataaacacactaccgctgcgacgaatatcgtggg
cactgttttcattggcatcggtctggcggagtccactgacttcatagacaaagacacggttagccgcctggg
tattggaatagcccaccaaagaagattgacctaacatggcgattctaaaaactcctgaacactaaagtttta
atttgctaattagcaaaatcatctacaaagacaaaaagagaagcgaccccaagttattccccaaagggaaaa
taaagtcaagccacttctcctcctgaactaaacagaaatctagctcaagctagccggggtaatactagcaac
cttaccccccaaacgattgatttgttgtaaggtctggttgagttgttcaaagggaacaattaccgccttatt
gctccgacgcactttgggataaccgggcgcagaaatggctgtaatttccacccggaacagcttacttgcttg
accaaagggaaccgtaccacccaacgcttttgctggggtattgccggcccgagaaggacgataggcccaacc
agcattactgccactggggcccacgatcgccgaaacagtattttgtcccaactccgttgccagacgactaga
actgcgctctaactgggaacggtcactgttagcataaccacggtaaagacggaacatgcgggtaaatcccac
agtgcgatcgcctacctgattattaaaacgataatagggaacaatattttccccaaagttctcttggtattc
gaccgaatcaatgtaagaatcgatatccgcatcaaaaccctgattttcatagagatcgagatggaaaatcac
ctcatcttcactgaaaggagcacgacccaacagatgtttgtagttgagttcaatgacccgggtttggaagtt
gttataaaggaatttcttcttataaagttccgacttggccacggtacggacaaattcccgcacactaataga
accattggtcaaaagagactcggcacccttgaggcgctcagaatccataatgtaatcgttgcctaaaacctg
acgatatacagcccgaatcaccattttggcatcatcaaggctaaagtcagggcgtaactccacaggacggga
ctcattatagggagccacccccaatctggaggctgcggttgtaattgccactaaataatctccttttactaa
ctaggttgacaaaaaatcctgatttctaggcaagggaaatcattgtgaccgtatctccttggcgattgatttg
ttgcaacttggcagacaaattatcgtaggaaactagatactccgctttaccacggcgcactcgggtgccacg
accgggggccgccccttgaatcacttgcaaacgatacatttggttacgactacccgccgaagttccccgcaa
actctcggccgtagaaccggcatacaccggagaagctgtattgcgagccaactcccgcgtcaatcgagaacg
ggaaccatttccctgggaacgatcactggtagcgtagccccgatacacctggaaagaacggctgaaaccaac
ggtttttttgattccgctgggtagcaaagccgcggaaataaggtacaacccaatctccaaagttttccgtgta
ttctacggagtcaatataggaattgatctccgcttcatagcccccctggtgataaagatcggtgtggaaagc
aatttccgactgatcgtagggagcccgtcccagtaaatgcttataattaagctcgataaaacgattttgtgg
gttggaatggaaaaacttctgccggtacacttccgacagagccacagcccgaacaaaatcccttaccgaaat
ctccctgccccggagcaaagattctgcactggtcagtcgttcctgggacatgagatgatcgttgcccaaaac
ttgacggtaaaccgccaggataacggcatccacctcttcctctgtactccgggaacgaagctcgaggggaat
agcctcatccacggccacaatgcccagacgctgggccgaaactaaactagtcatttgttctcctagataag
tgtcaataaatacgacaagacctggaaaagcggaatattgtcaaccagttgttaacactttcccacccttcc
aggccagctagaattaaaaactgactagctcagagcattgatggcgtaatcgaggtaggaattagcttcgtc
acgggcatcgccactcaagccgtggttagctttgatgtatttcagagcttcaacataccagctgggggagag
gtcaaaggtgcggttgatttcatcaataccggcgatcaagtactcatccaaaggaccggtaccaccagcaac
taagcagtaggtaacgatgcggaggtagtagccgatgtcccgggcacacttgtctttaccccgttgatccgc
agcaaagttgttgccttgggtttgggtggtgtaggggaatttgttataaacggcttgggcagcaccatttac

caagctctgggcattgtcggtcagagctttagcggcttgcaaaccagcattagcttgacgtagacgaccgaa
agcagcgaccttgagagtcagcggtggaaacggcttcagttaaaggggttttcatatttgcaattcggtgct
gctcagaaatgttctggtatctcccgttagaatgtgaaaattttcctagatatgtcagctttaagctggatt
ttctcgggtatggaacctgagtaaagttggttttaaaataaccagactaggctacggcagcagcggcgcgg
tcgaagtaaccagcgatttcagcaacgatagcactgcaatcaccacgggtgatgccattgggatcgttaacg
atgtccagggcagcttctttcatttttgaacgccagcagctacggaagcaccgggaacacccagggcaacg
taggtttcacggagaccgttcaagcaacgatcttctagaacggaagcgtcgccggtgaaggttgcgtaggta
acatagcggaggatgatttccatgtcacgcaaacaagcagccatacgacggctggtgtaggcgtttccaccg
ggttggattaattggggctgttcggcgaacaaagcacgagcagcgttggaaacgatagcggaagcattaccg
gtgatgcggttaacagaatcaatccgtttgttgccttcagcaacggtagcgctcaaagcatctaactgagaa
ccagagaggtactcgccgcgagcatcagcttgggaaacaacccgagtgaatacgtcgaacattgaattaatc
tcctacttgactttatgagttgggattttcttaaacacaattcccccggataaactgagggagtccaaagta
atgaccctagagttattgttactgatctccattaactttcgttaactacccggggatttatgagagatatta
cctaaataaatccagggagaaacacggaggcagcgacaagggccaccgggatgctcaaacagctcagcgcct
aggcttgaatgcttttgcaatcccacagttaactttatacaacggtgatgggacttatgtctgttacatctt
gttaattttattcctgcttttttgttaagtaatgttgcaggggattctcagattgtcctggattgggaaggg
aagacaaccagtttcgttcagcttatgttttagggctaaaattatgcaatt

Appendix 9: Sequence analysis of the pLAH.nrsB.ble expression vector.

Key: *nrsB* promoter + regulatory elements (red), ble selectable marker (blue).

```
Query  91    CGCCACCACA-TATTCAT-TGATCCTCATC-AGATCCCACAG-TG-TGCCGAATC-TGCT  144
             |||||||||| || ||||  | ||||||||| |||||||||| || ||||| ||| ||||
Sbjct  284   CGCCACCACATTACTCATGGGCTCCTCATCAAGATCCCACAGTTGTTGCCGGATCTTGCT  343

Query  145   A-CGGAAATGATCCGCTCTGGGTTTT-GCATCAGATA-TGAAAAATTGGAAAT-CTC-TAC  199
             | ||||||||||||||||||||||| |||||||||| ||||||||| ||||| ||| |||
Sbjct  344   ACCGGAAATGATCCGCTCTGGGTTTTGCATCAGATATTGAAAAATTTGAAATTCTCTTAC  403

Query  200   -GCTAAAGC-A-TT-CTGTCTTTCTAGGTTTAGTGGCTCCGAGATAGTTACCGATAACAG  255
              | |||||| | || ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  404   GGTTAAAGCAATTTCCTGTCTTTCTAGGTTTAGTGGCTCCGAGATAGTTACCGATAACAG  463

Query  256   ATTATTACTGGGATCAAGGCTGAAGTTGCCCAAAGTTAAAATTTGCGGTTGGAATTGTGG  315
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  464   ATTATTACTGGGATCAAGGCTGAAGTTGCCCAAAGTTAAAATTTGCGGTTGGAATTGTGG  523

Query  316   CGATCGCCGTTGTAGTGCCCGCAGTCTTGCTAATAGCTCTGCCATCACAAACGGTTTTGT  375
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  524   CGATCGCCGTTGTAGTGCCCGCAGTCTTGCTAATAGCTCTGCCATCACAAACGGTTTTGT  583

Query  376   TAGATAGTCATCTGCCCCGGCATCTAGTCCTTCGACACGGTTTTCCGGTTCTCCTAACGC  435
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  584   TAGATAGTCATCTGCCCCGGCATCTAGTCCTTCGACACGGTTTTCCGGTTCTCCTAACGC  643

Query  436   TGTTAACATCAACACCGGCAAGGAATTACCCTGGGTTCTCAGTTTTTGACAGAGTTCCAA  495
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  644   TGTTAACATCAACACCGGCAAGGAATTACCCTGGGTTCTCAGTTTTTGACAGAGTTCCAA  703

Query  496   ACCCGATAATCCCGGCAGTAACCAATCCACAATGGCAAGGGTGTATTCCGTCCATTGATT  555
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  704   ACCCGATAATCCCGGCAGTAACCAATCCACAATGGCAAGGGTGTATTCCGTCCATTGATT  763

Query  556   TTCCAAATAATCCCAAGCTTGGGAGCCATCCGTCACCCAATCCACCACATACTTTTCACT  615
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  764   TTCCAAATAATCCCAAGCTTGGGAGCCATCCGTCACCCAATCCACCACATACTTTTCACT  823

Query  616   AACTAGCACTTTCTTAATAGCCATTCCCAAATCCGTCTCATCTTCCACCAGCAAAATTCG  675
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  824   AACTAGCACTTTCTTAATAGCCATTCCCAAATCCGTCTCATCTTCCACCAGCAAAATTCG  883

Query  676   CATCGCCTCTGCCTTTTTTATAACGGTCTGATCTTAGCGGGGGAAGGAGATTTTCACCTG  735
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  884   CATCGCCTCTGCCTTTTTTATAACGGTCTGATCTTAGCGGGGGAAGGAGATTTTCACCTG  943

Query  736   AATTTCATACCCCCTTTGGCAGACTGGGAAAATCTTGGACAAATTCCCAATTTGAGGTCA  795
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  944   AATTTCATACCCCCTTTGGCAGACTGGGAAAATCTTGGACAAATTCCCAATTTGAGGTCA  1003

Query  796   TATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGACGTCGCCGGAGCGGT  855
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1004  TATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGACGTCGCCGGAGCGGT  1063

Query  856   CGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGG  915
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1064  CGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGG  1123

Query  916   TGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGA  975
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1124  TGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGA  1183

Query  976   CAACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGA  1035
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1184  CAACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGA  1243

Query  1036  GGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCA  1095
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1244  GGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCA  1303

Query  1096  GCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGC  1155
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1304  GCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGC  1363

Query  1156  CGAGGAGCAGGACTGAGG  1173
             ||||||||||||||||||
Sbjct  1364  CGAGGAGCAGGACTGAGG  1381
```

Appendix 10: Sequenced analysis of pLAH.A2.GES.F expression vector

- **Key:** Primers used for sequencing (red), Match with *psbA2* gene (blue)
- Sequence results show 99% match with the psbA2 gene but no match with GES gene.
- Identities = 203/204 (99%), Gaps = 1/204 (0%), Strand=Plus/Plus

```
Query  12   ACTCTTCTC-ACCCCCAAAACGCCCTCTGTTTACCCATGGAAAAAACGACAATTACAAGA  70
            ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  42   ACTCTTCTCAACCCCCAAAACGCCCTCTGTTTACCCATGGAAAAAACGACAATTACAAGA  101

Query  71   AAGTAAAACTTATGTCATCTATAAGCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTA  130
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  102  AAGTAAAACTTATGTCATCTATAAGCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTA  161

Query  131  CAAAACTCTCATTAATCCTTTAGACTAAGTTTAGTCAGTTCCAATCTGAACATCGACAAA  190
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162  CAAAACTCTCATTAATCCTTTAGACTAAGTTTAGTCAGTTCCAATCTGAACATCGACAAA  221

Query  191  TACATAAGGAATTATAACCATATG  214
            |||||||||||||||||||||||
Sbjct  222  TACATAAGGAATTATAACCATATG  245
```

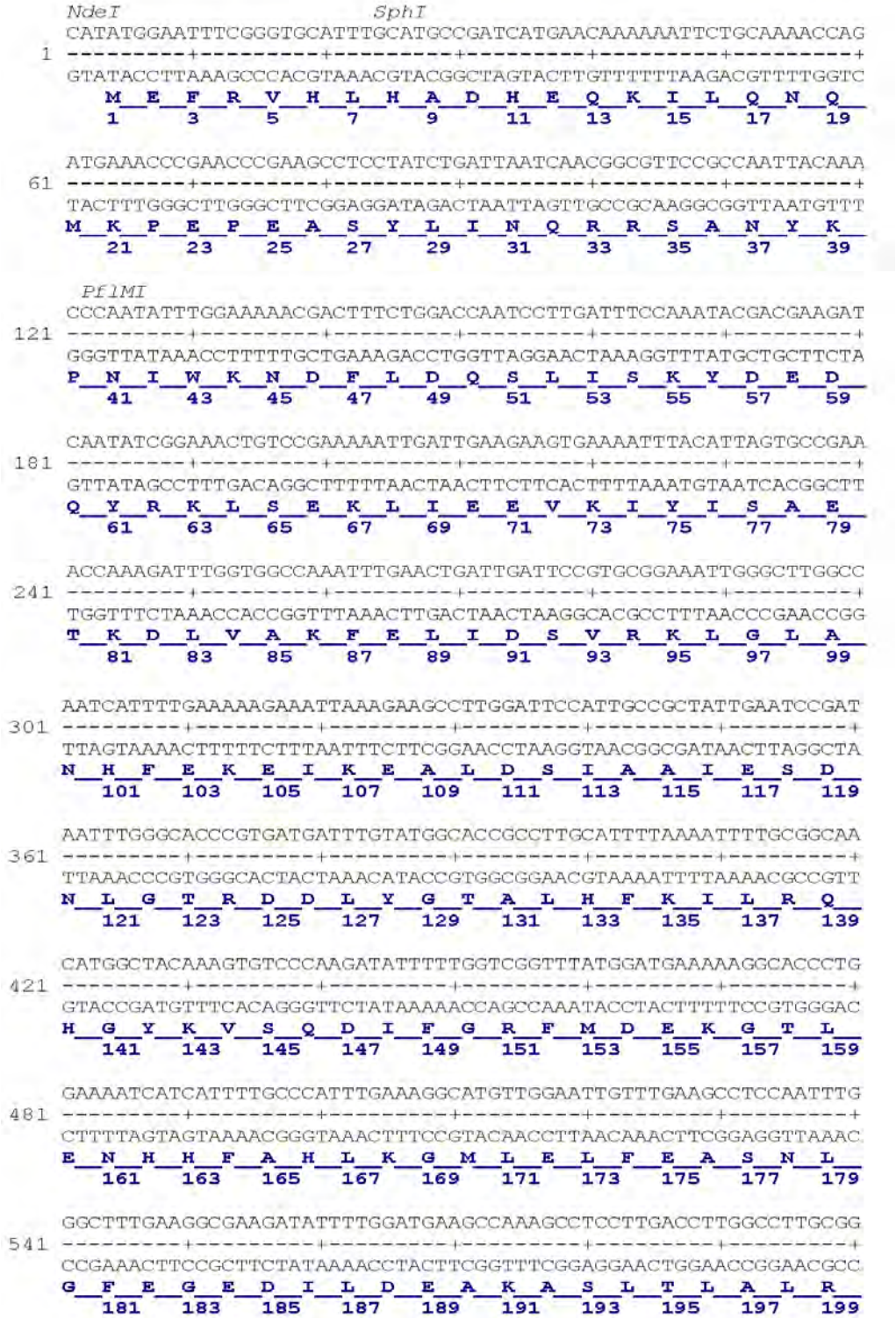
- **Visual of GES under the psbA2 promoter**

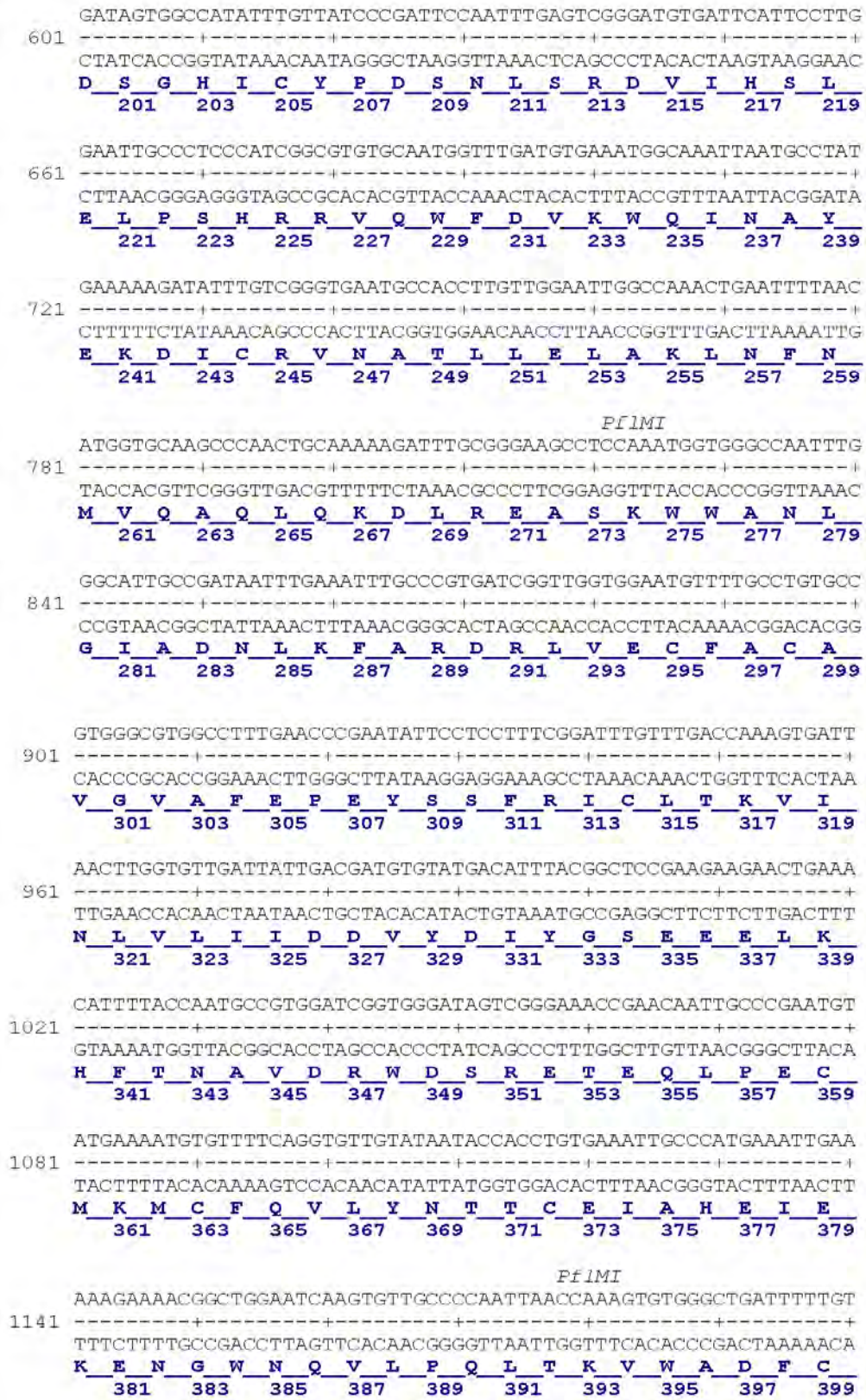
Sequence alignment results presented above show a match only to the psbA2 promoter (BLUE).

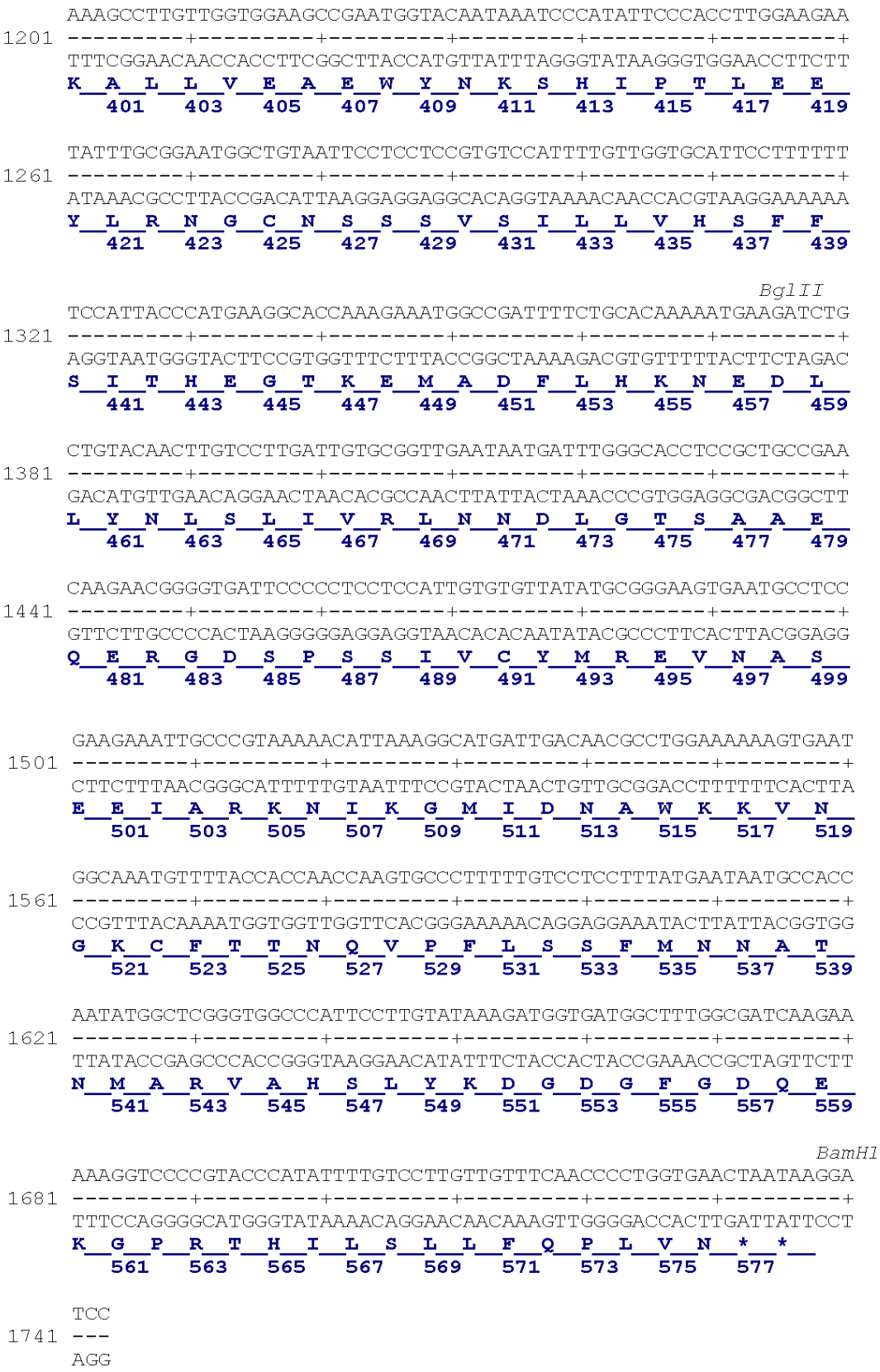
```
cccagaactatggtaaaggcgcacgaaaaaccgccaggtaaactcttctcaacccccaaaacgccctctgtttacccatggaaaaaacga
caattacaagaaagtaaaacttatgtcatctataagcttcgtgtatattaacttcctgttacaaagctttacaaaactctcattaatcct
ttagactaagtttagtcagttccaatctgaacatcgacaaatacataaggaattataaccaTATgCGCCCACGCTTCTCTG CTTG
CACGCCTTTGGCATCGGCGA TGCCTCTAAG TTCAACTCCT CTCATCAACG GGGATAACTC TCAGCGTAAAAACACACGTC
AACACATGGA GGAGAGCAGC AGCAAGAGGA GAGAATATCT GCTGGAGGAAACGACGCGAA AACTGCAGAG AAACGACACC
GAATCGGTGG AGAAACTCAA GCTTATCGACAACATCCAAC AGTTGGGAAT CGGCTACTAT TTTGAGGACG CCATCAACGC
CGTACTCCGCTCGCCTTTCT CCACCGGAGA AGAAGACCTC TTCACCGCTG CTCTGCGCTT CCGCTTGCTCCGCCACAACG
GCATCGAAAT CAGCCCTGAA ATATTCCTAA AATTCAAGGA CGAGAGGGGAAAATTCGACG AATCGGACAC GCTAGGGTTA
CTGAGCTTGT ACGAAGCGTC AAATTTGGGGGTTGCAGGAG AAGAAATATT GGAGGAGGCT ATGGAGTTTG CGGAGGCTCG
CCTGAGACGGTCGCTGTCAG AGCCGGCGGC GCCGCTTCAT GGTGAGGTGG CGCAAGCGCT AGATGTGCCGAGGCATCTGA
GAATGGCGAG GTTGGAAGCG AGACGATTCA TCGAGCAGTA TGGTAAACAGAGCGATCATG ATGGAGATCT TTTGGAGCTG
GCAATTTTGG ATTATAATCA AGTTCAGGCTCAACACCAAT CCGAACTCAC TGAAATAATC AGGTGGTGGA AGGAGCTCGG
TTTGGTGGATAAGTTGAGTT TTGGGCGAGA CAGACCATTG GAGTGCTTTT TGTGGACCGT GGGGCTCCTCCCAGAGCCCA
AGTATTCGAG CGTTAGAATA GAGTTGGCGA AAGCCATCTC TATTCTCTTAGTGATCGATG ATATTTTCGA TACCTATGGA
GAGATGGATG ACCTCATCCT CTTCACCGATGCAATTCGAA GATGGGATCT TGAAGCAATG GAGGGGCTCC CTGAGTACAT
GAAAATATGCTACATGGCGT TGTACAATAC CACCAATGAA GTATGCTACA AAGTGCTCAG GGATACTGGACGGATTGTCC
TCCTTAACCT CAAATCTACG TGGATAGACA TGATTGAAGG TTTCATGGAGGAAGCAAAAT GGTTCAATGG TGGAAGTGCA
CCAAAATTGG AAGAGTATAT AGAGAATGGAGTGTCCACGG CAGGAGCATA CATGGCTTTT GCACACATCT TCTTTCTCAT
AGGAGAAGGTGTTACACACC AAAATTCCCA ACTCTTCACC CAAAAACCCT ACCCCAAGGT CTTCTCCGCCGCCGGCCGCA
TTCTTCGCCT CTGGGATGAT CTCGGAACCG CCAAGGAAGA GCAAGAGCGAGGAGATCTGG CTTCGTGCGT GCAGTTATTT
ATGAAAGAGA AGTCGTTGAC GGAAGAGGAGGCAAGAAGTC GCATTTTGGA AGAGATAAAA GGATTATGGA GGGATCTGAA
TGGGGAACTGGTCTACAACA AGAATTTGCC GTTATCCATA ATCAAAGTCG CACTTAACAT GGCGAGAGCTTCTCAAGTTG
TGTACAAGCA CGATCAAGAC ACTTATTTTT CAAGCGTAGA
CAATTATGTGGATGGTAGgatccccgacctgcagggggggggggcgctgaggtctgcctcgtgaagaaggtgtt
```

No match for GES gene

Appendix 11: Condon optimised α-farnesene synthase gene and its corresponding protein sequence.

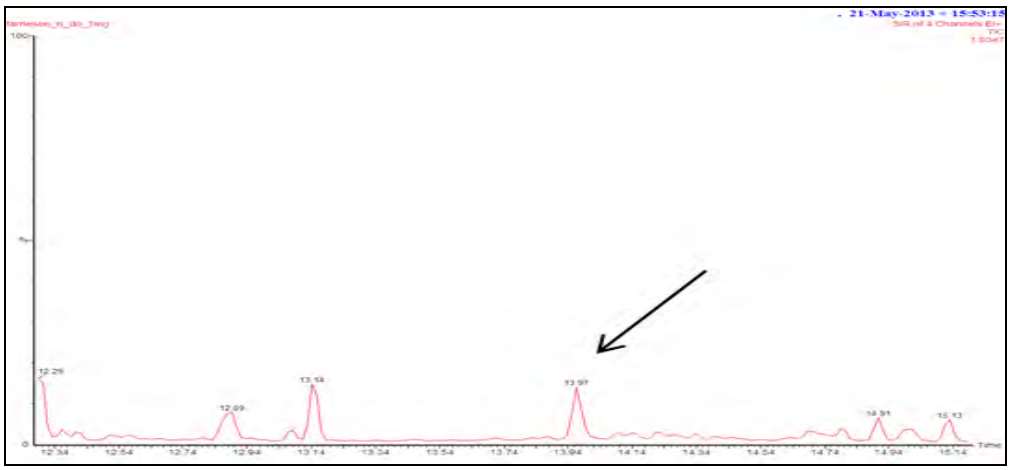


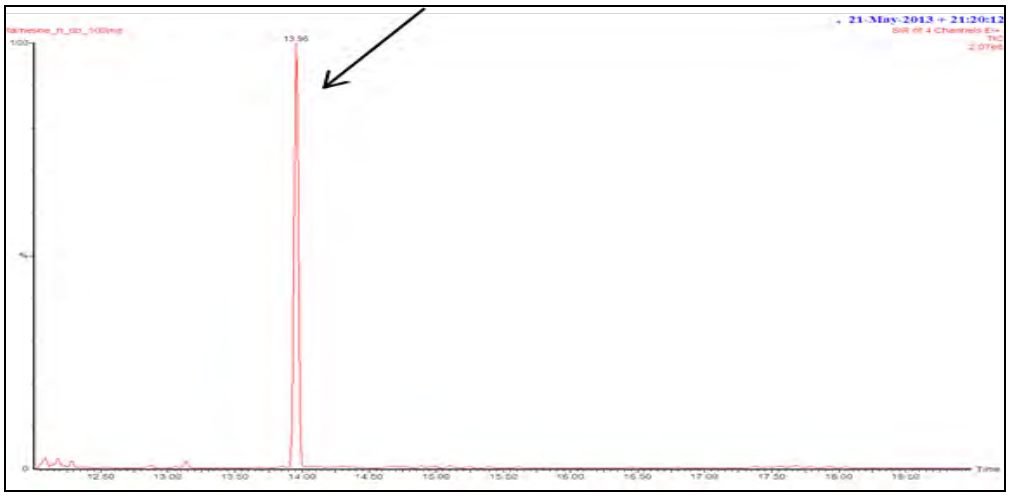

GATAGTGGCCATATTTGTTATCCCGATTCCAATTTGAGTCGGGATGTGATTCATTCCTTG
601
CTATCACCGGTATAAACAATAGGGCTAAGGTTAAACTCAGCCCTACACTAAGTAAGGAAC
D S G H I C Y P D S N L S R D V I H S L
201 203 205 207 209 211 213 215 217 219
GAATTGCCCTCCCATCGGCGTGTGCAATGGTTTGATGTGAAATGGCAAATTAATGCCTAT
661
CTTAACGGGAGGGTAGCCGCACACGTTACCAAACTACACTTTACCGTTTAATTACGGATA
E L P S H R R V Q W F D V K W Q I N A Y
221 223 225 227 229 231 233 235 237 239
GAAAAAGATATTTGTCGGGTGAATGCCACCTTGTTGGAATTGGCCAAACTGAATTTTAAC
721
CTTTTTCTATAAACAGCCCACTTACGGTGGAACAACCTTAACCGGTTTGACTTAAAATTG
E K D I C R V N A T L L E L A K L N F N
241 243 245 247 249 251 253 255 257 259
PflMI
ATGGTGCAAGCCCAACTGCAAAAAGATTTGCGGGAAGCCTCCAAATGGTGGGCCAATTTG
781
TACCACGTTCGGGTTGACGTTTTTCTAAACGCCCTTCGGAGGTTTACCACCCGGTTAAAC
M V Q A Q L Q K D L R E A S K W W A N L
261 263 265 267 269 271 273 275 277 279
GGCATTGCCGATAATTTGAAATTTGCCCGTGATCGGTTGGTGGAATGTTTTGCCTGTGCC
841
CCGTAACGGCTATTAAACTTTAAACGGGCACTAGCCAACCACCTTACAAAACGGACACGG
G I A D N L K F A R D R L V E C F A C A
281 283 285 287 289 291 293 295 297 299
GTGGGCGTGGCCTTTGAACCCGAATATTCCTCCTTTCGGATTTGTTTGACCAAAGTGATT
901
CACCCGCACCGGAAACTTGGGCTTATAAGGAGGAAAGCCTAAACAAACTGGTTTCACTAA
V G V A F E P E Y S S F R I C L T K V I
301 303 305 307 309 311 313 315 317 319
AACTTGGTGTTGATTATTGACGATGTGTATGACATTTACGGCTCCGAAGAAGAACTGAAA
961
TTGAACCACAACTAATAACTGCTACACATACTGTAAATGCCGAGGCTTCTTCTTGACTTT
N L V L I I D D V Y D I Y G S E E E L K
321 323 325 327 329 331 333 335 337 339
CATTTTACCAATGCCGTGGATCGGTGGGATAGTCGGGAAACCGAACAATTGCCCGAATGT
1021
GTAAAATGGTTACGGCACCTAGCCACCCTATCAGCCCTTTGGCTTGTTAACGGGCTTACA
H F T N A V D R W D S R E T E Q L P E C
341 343 345 347 349 351 353 355 357 359
ATGAAAATGTGTTTTCAGGTGTTGTATAATACCACCTGTGAAATTGCCCATGAAATTGAA
1081
TACTTTTACACAAAAGTCCACAACATATTATGGTGGACACTTTAACGGGTACTTTAACTT
M K M C F Q V L Y N T T C E I A H E I E
361 363 365 367 369 371 373 375 377 379
PflMI
AAAGAAAACGGCTGGAATCAAGTGTTGCCCCAATTAACCAAAGTGTGGGCTGATTTTTGT
1141
TTTCTTTTGCCGACCTTAGTTCACAACGGGGTTAATTGGTTTCACACCCGACTAAAAACA
K E N G W N Q V L P Q L T K V W A D F C
381 383 385 387 389 391 393 395 397 399

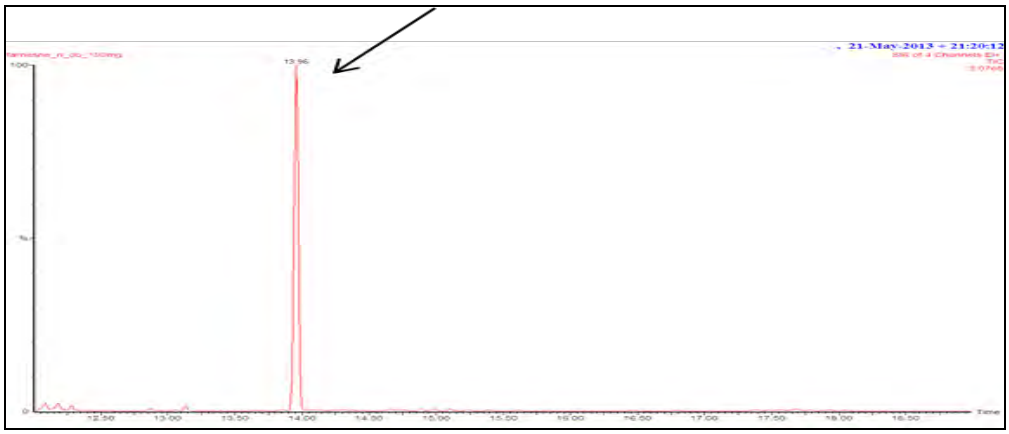

AAAGCCTTGTTGGTGGAAGCCGAATGGTACAATAAATCCCATATTCCCACCTTGGAAGAA
1201
TTTCGGAACAACCACCTTCGGCTTACCATGTTATTTAGGGTATAAGGGTGGAACCTTCTT
K A L L V E A E W Y N K S H I P T L E E
401 403 405 407 409 411 413 415 417 419
TATTTGCGGAATGGCTGTAATTCCTCCTCCGTGTCCATTTTGTTGGTGCATTCCTTTTTT
1261
ATAAACGCCTTACCGACATTAAGGAGGAGGCACAGGTAAAACAACCACGTAAGGAAAAAA
Y L R N G C N S S S V S I L L V H S F F
421 423 425 427 429 431 433 435 437 439
BglII
TCCATTACCCATGAAGGCACCAAAGAAATGGCCGATTTTCTGCACAAAAATGAAGATCTG
1321
AGGTAATGGGTACTTCCGTGGTTTCTTTACCGGCTAAAAGACGTGTTTTTACTTCTAGAC
S I T H E G T K E M A D F L H K N E D L
441 443 445 447 449 451 453 455 457 459
CTGTACAACTTGTCCTTGATTGTGCGGTTGAATAATGATTTGGGCACCTCCGCTGCCGAA
1381
GACATGTTGAACAGGAACTAACACGCCAACTTATTACTAAACCCGTGGAGGCGACGGCTT
L Y N L S L I V R L N N D L G T S A A E
461 463 465 467 469 471 473 475 477 479
CAAGAACGGGGTGATTCCCCCTCCTCCATTGTGTGTTATATGCGGGAAGTGAATGCCTCC
1441
GTTCTTGCCCCACTAAGGGGGAGGAGGTAACACACAATATACGCCCTTCACTTACGGAGG
Q E R G D S P S S I V C Y M R E V N A S
481 483 485 487 489 491 493 495 497 499
GAAGAAATTGCCCGTAAAAACATTAAAGGCATGATTGACAACGCCTGGAAAAAAGTGAAT
1501
CTTCTTTAACGGGCATTTTTGTAATTTCCGTACTAACTGTTGCGGACCTTTTTTCACTTA
E E I A R K N I K G M I D N A W K K V N
501 503 505 507 509 511 513 515 517 519
GGCAAATGTTTTACCACCAACCAAGTGCCCTTTTTGTCCTCCTTTATGAATAATGCCACC
1561
CCGTTTACAAAATGGTGGTTGGTTCACGGGAAAAACAGGAGGAAATACTTATTACGGTGG
G K C F T T N Q V P F L S S F M N N A T
521 523 525 527 529 531 533 535 537 539
AATATGGCTCGGGTGGCCCATTCCTTGTATAAAGATGGTGATGGCTTTGGCGATCAAGAA
1621
TTATACCGAGCCCACCGGGTAAGGAACATATTTCTACCACTACCGAAACCGCTAGTTCTT
N M A R V A H S L Y K D G D G F G D Q E
541 543 545 547 549 551 553 555 557 559
BamH1
AAAGGTCCCCGTACCCATATTTTGTCCTTGTTGTTTCAACCCCTGGTGAACTAATAAGGA
1681
TTTCCAGGGGCATGGGTATAAAACAGGAACAACAAAGTTGGGGACCACTTGATTATTCCT
K G P R T H I L S L L F Q P L V N * *
561 563 565 567 569 571 573 575 577
TCC
1741 ---
AGG

Appendix 12: GC-MS results of Farnesene in n-dodecane.

(A) GC-MS results of farnesene in n-dodecane standards

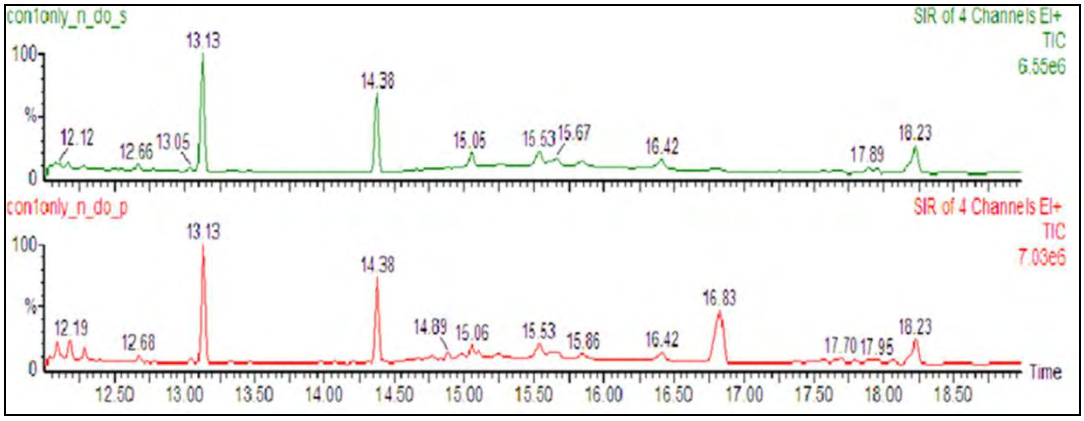




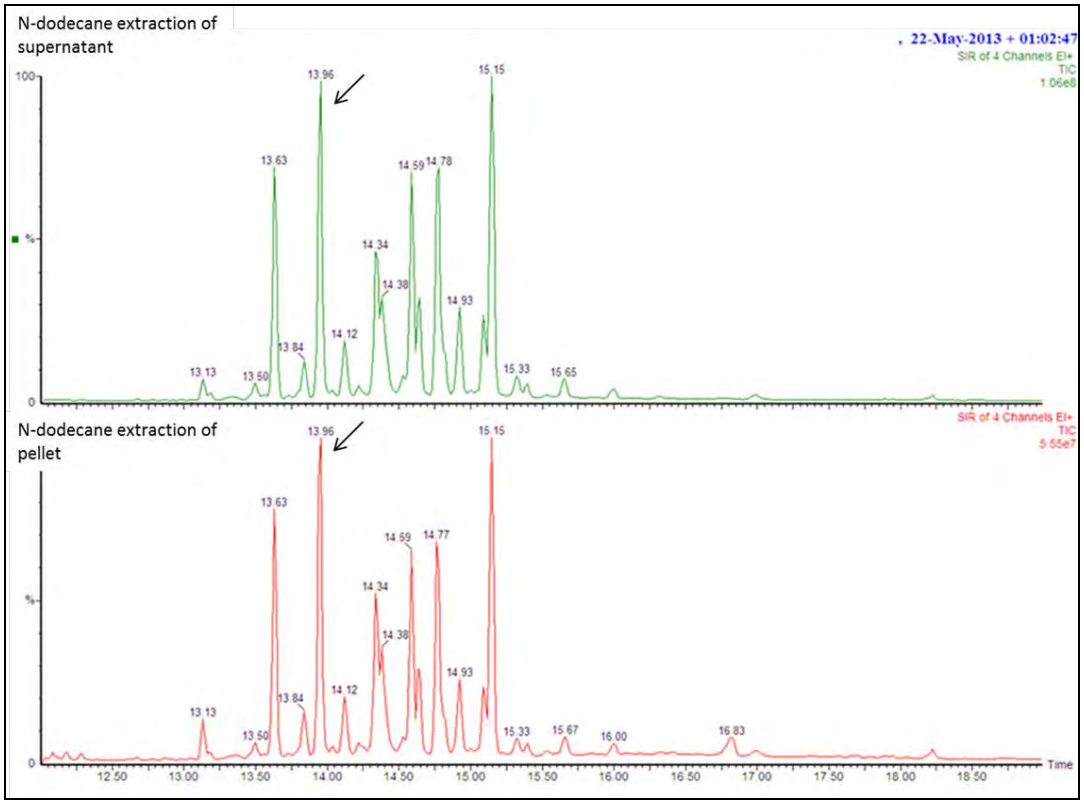


(B) GC-MS results of farnesene in n-dodecane controls

- Negative controls

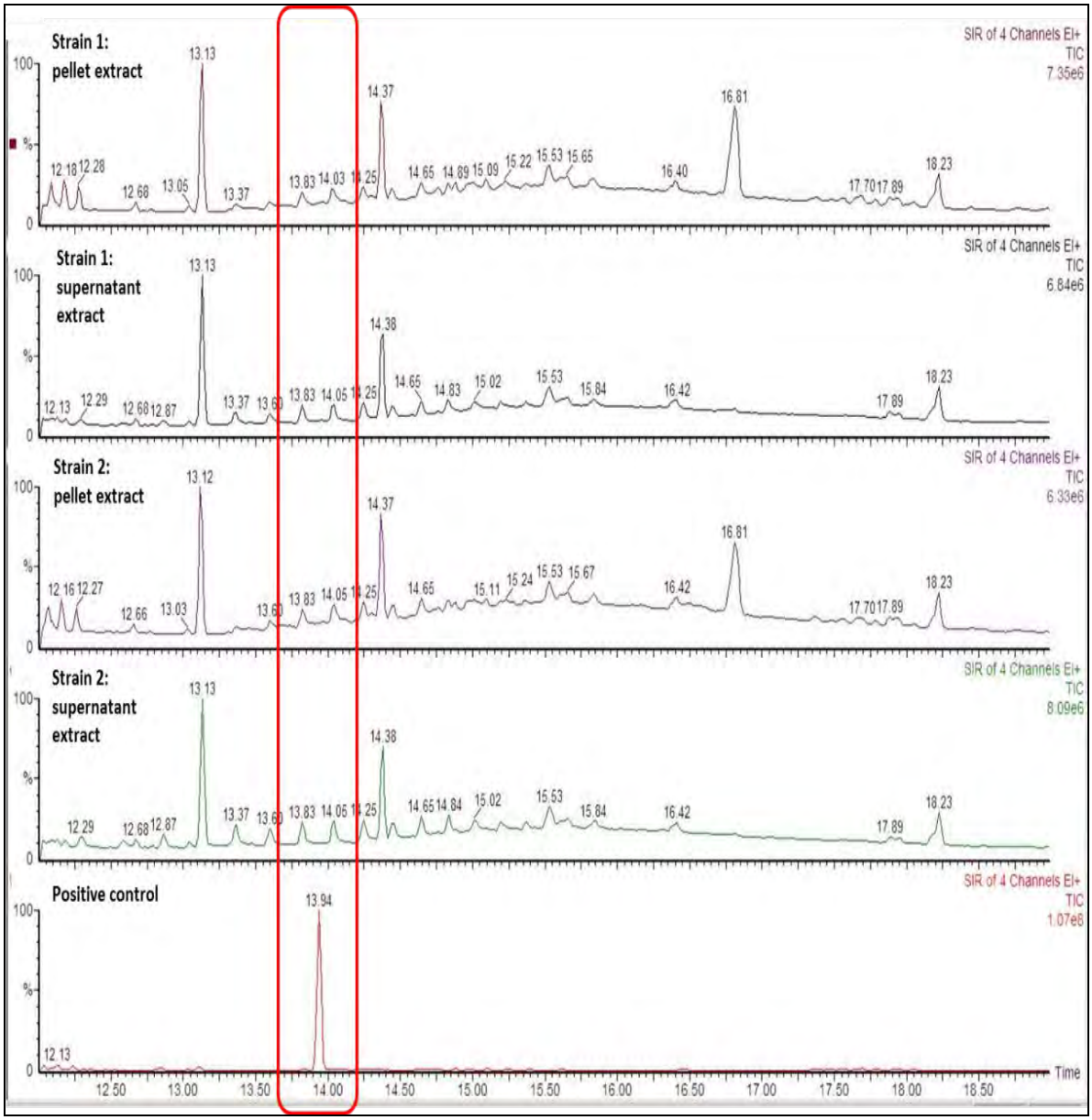


- Positive controls: cultures spiked with 0.02% farnesene



(C) GC-MS results of samples extracted in n-dodecane

GC-MS results of n-dodecane supernatant and cell extracts of *Synechocystis* strain one and strain two expressing farnesene synthase under the *psbA2* and nrsBR promoter respectively. No peak corresponding to α-farnesene observed at RT~ 13.95 min. Positive control is a 50 ppm standard of farnesene in n-dodecane.



Appendix 13: pLAH.nrsB.GES.HA sequence results

Key: Primer used for seq (blue), deleted area from con2.GES plasmid upstream nrsB promoter (red), full nrsB promoter & regulator element (purple), GES gene (orange) showing start codon (green), A-tag (underlined) and stop codon (bold and stared*).

- **Forward reaction: shows perfect match with all of the nrsBR + front part of GES.**

```
Seq.result.GE      1 ------------------------------------------------      0

visual             1 tgtgtcgcccctctacacagcccagaactatggtaaaggcgcacgaaaaa     50

Seq.result.GE      1 CGGCCCGCAGT--ACTCTTCTC-ACCCCCAAAACGCCCTCTGTTTACCCA     47
                     |.|||.|  ||  ||||||||| |||||||||||||||||||||||||||
visual            51 ccgccag--gtaaactcttctcaacccccaaaacgccctctgtttaccca     98

Seq.result.GE     48 TGGAAAAAACGACAATTACAAGAAAGTAAAACTTATGTCATCTAT-----     92
                     ||||||||||||||||||||||||||||||||||||||||||||||
visual            99 tggaaaaaacgacaattacaagaaagtaaaacttatgtcatctatcagga    148

Seq.result.GE     93 --------------AAGCTTGCATGCGACTACGGGCAAAGAGGCGACGGG    128
                                 |||||||||||||||||||||||||||||||||||||
visual           149 agttaatatacacgaagcttgcatgcgactacgggcaaagaggcgacggg    198

Seq.result.GE    129 TATTCATGGCGATAGGGTGAACCGATAGCCTTGACCGGGAACTGTTTTAA    178
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
visual           199 tattcatggcgatagggtgaaccgatagccttgaccgggaactgttttaa    248

Seq.result.GE    179 TTGGGCAAGGACAATTTTGTTGAGCTAGCTTGCGTCGTATCAAACGCATT    228
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
visual           249 ttgggcaaggacaattttgttgagctagcttgcgtcgtatcaaacgcatt    298

Seq.result.GE    229 TGGGCCGCCACCACATTACTCATGGGCTCCTCATCAAGATCCCACAGTTG    278
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
visual           299 tgggccgccaccacattactcatgggctcctcatcaagatcccacagttg    348

Seq.result.GE    279 TTGCCGGATCTTGCTACCGGAAATGATCCGCTCTGGGTTTTGCATCAGAT    328
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
visual           349 ttgccggatcttgctaccggaaatgatccgctctgggttttgcatcagat    398

Seq.result.GE    329 ATTGAAAAATTTGAAATTCTCTTACGGTTAAAGCAATTTCCTGTCTTTCT    378
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
visual           399 attgaaaaatttgaaattctcttacggttaaagcaatttcctgtctttct    448

Seq.result.GE    379 AGGTTTAGTGGCTCCGAGATAGTTACCGATAACAGATTATTACTGGGATC    428
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
visual           449 aggtttagtggctccgagatagttaccgataacagattattactgggatc    498

Seq.result.GE    429 AAGGCTGAAGTTGCCCAAAGTTAAAATTTGCGGTTGGAATTGTGGCGATC    478
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
visual           499 aaggctgaagttgcccaaagttaaaatttgcggttggaattgtggcgatc    548

Seq.result.GE    479 GCCGTTGTAGTGCCCGCAGTCTTGCTAATAGCTCTGCCATCACAAACGGT    528
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
visual           549 gccgttgtagtgcccgcagtcttgctaatagctctgccatcacaaacggt    598

Seq.result.GE    529 TTTGTTAGATAGTCATCTGCCCCGGCATCTAGTCCTTCGACACGGTTTTC    578
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
visual           599 tttgttagatagtcatctgccccggcatctagtccttcgacacggttttc    648

Seq.result.GE    579 CGGTTCTCCTAACGCTGTTAACATCAACACCGGCAAGGAATTACCCTGGG    628
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
visual           649 cggttctcctaacgctgttaacatcaacaccggcaaggaattaccctggg    698

Seq.result.GE    629 TTCTCAGTTTTTGACAGAGTTCCAAACCCGATAATCCCGGCAGTAACCAA    678
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
visual           699 ttctcagtttttgacagagttccaaacccgataatcccggcagtaaccaa    748

Seq.result.GE    679 TCCACAATGGCAAGGGTGTATTCCGTCCATTGATTTTCCAAATAATCCCA    728
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
visual           749 tccacaatggcaagggtgtattccgtccattgattttccaaataatcccca    798
```

```
Seq.result.GE    729 AGCTTGGGAGCCATCCGTCACCCAATCCACCACATACTTTTCACTAACTA    778
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual           799 agcttgggagccatccgtcacccaatccaccacatactttcactaacta    848

Seq.result.GE    779 GCACTTTCTTAATAGCCATTCCCAAATCCGTCTCATCTTCCACCAGCAAA    828
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual           849 gcactttcttaatagccattcccaaatccgtctcatcttccaccagcaaa    898

Seq.result.GE    829 ATTCGCATCGCCTCTGCCTTTTTTATAACGGTCTGATCTTAGCGGGGGAA    878
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual           899 attcgcatcgcctctgccttttttataacggtctgatcttagcgggggaa    948

Seq.result.GE    879 GGAGATTTTCACCTGAATTTCATACCCCCTTTGGCAGACTGGGAAAATCT    928
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual           949 ggagattttcacctgaatttcataccccctttggcagactgggaaaatct    998

Seq.result.GE    929 TGGACAAATTCCCAATTTGAGGTCATATGCGCCCACGCTTCTCTGCTTGC    978
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual           999 tggacaaattcccaatttgaggtCATATgCGCCCACGCTTCTCTGCTTGC   1048

Seq.result.GE    979 ACGCCTTTGGCATCGGCGATGCCTCTAAGTTCAACTCCTCTCATCAACGG   1028
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual          1049 ACGCCTTTGGCATCGGCGATGCCTCTAAGTTCAACTCCTCTCATCAACGG   1098

Seq.result.GE   1029 GGATAACTCTCAGCGTAAAAACACACGTCAACACATGGAGGAGAGCAGCA   1078
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual          1099 GGATAACTCTCAGCGTAAAAACACACGTCAACACATGGAGGAGAGCAGCA   1148

Seq.result.GE   1079 GCA----GAGAGAGATATCTGCTG--AGAAACGACGCG-AAACTGCAGAG   1121
                     |||    |||||| ||||||||||  .|||||||||||| ||||||||||||
visual          1149 GCAAGAGGAGAGA-ATATCTGCTGGAGGAAACGACGCGAAAACTGCAGAG   1197
```

- **Reverse reaction:** shows a very good match (for about 1000 bp) with GES.

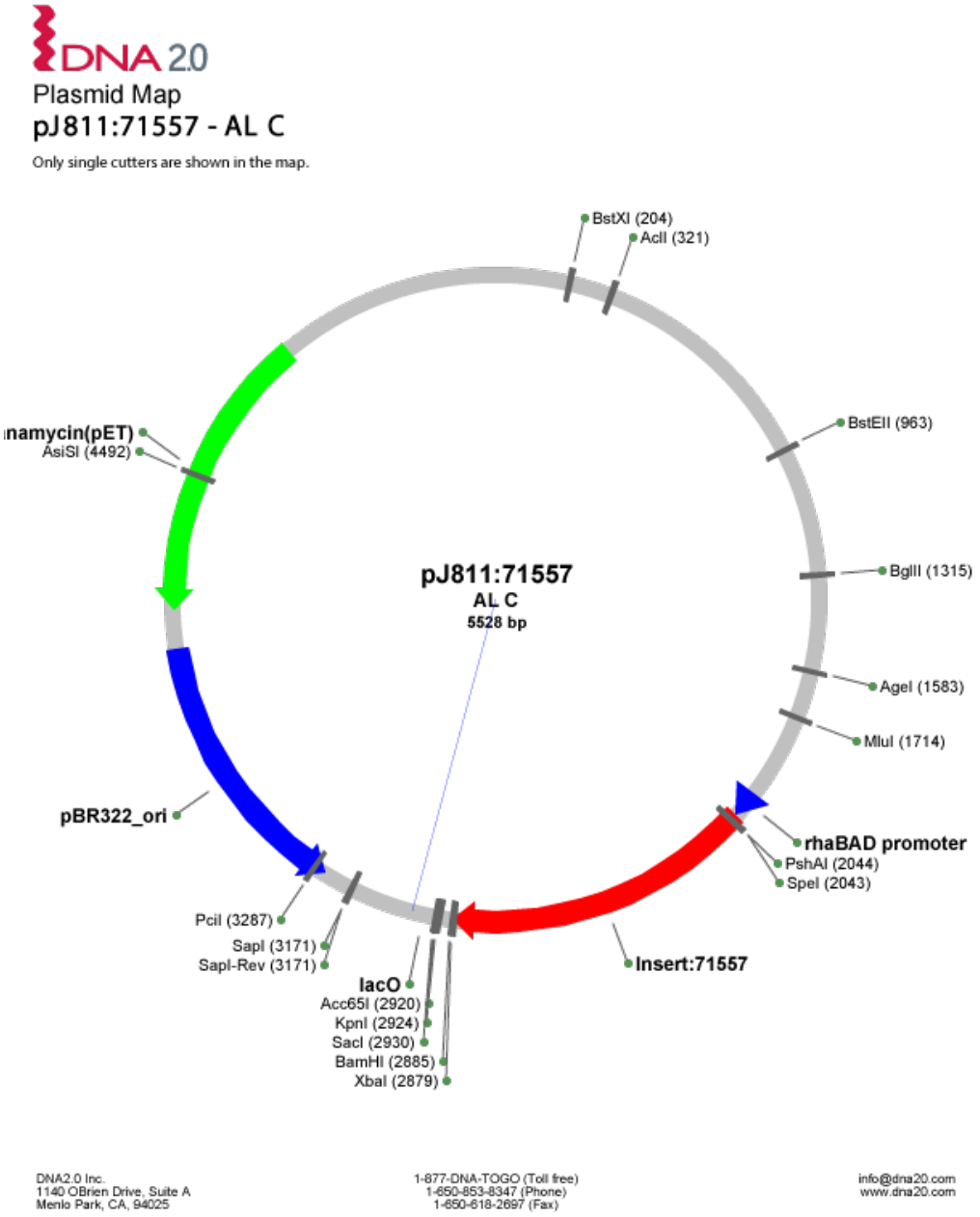
```
LA_pLAH_nrsB_    201                    CAAGTTCAGGCTACTGAAATAATCAGGT      250
                                        ||||||||||||||||||||||||||||
visual             1                    CAATCCGAACTCACTGAAATAATCAGGT       34

LA_pLAH_nrsB_    251 GGTGGAAAGGAGCTCGGTTTGGTGGATAAGTTGAGTTTTGGGCGAGACAG    300
                     ||||| ||||||||||||||||||||||||||||||||||||||||||||||
visual            35 GGTGG-AAGGAGCTCGGTTTGGTGGATAAGTTGAGTTTTGGGCGAGACAG     83

LA_pLAH_nrsB_    301 ACCATTGGAGTGCTTTTTGTGGACCGTGGGGCTCCTCCCAGAGCCCAAGT    350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual            84 ACCATTGGAGTGCTTTTTGTGGACCGTGGGGCTCCTCCCAGAGCCCAAGT    133

LA_pLAH_nrsB_    351 ATTCGAGCGTTAGAATAGAGTTGGCGAAAGCCATCTCTATTCTCTTAGTG    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual           134 ATTCGAGCGTTAGAATAGAGTTGGCGAAAGCCATCTCTATTCTCTTAGTG    183

LA_pLAH_nrsB_    401 ATCGATGATATTTTCGATACCTATGGAGAGATGGATGACCTCATCCTCTT    450
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual           184 ATCGATGATATTTTCGATACCTATGGAGAGATGGATGACCTCATCCTCTT    233


LA_pLAH_nrsB_    451 CACCGATGCAATTCGAAGATGGGATCTTGAAGCAATGGAGGGGCTCCCTG    500
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual           234 CACCGATGCAATTCGAAGATGGGATCTTGAAGCAATGGAGGGGCTCCCTG    283

LA_pLAH_nrsB_    501 AGTACATGAAAATATGCTACATGGCGTTGTACAATACCACCAATGAAGTA    550
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual           284 AGTACATGAAAATATGCTACATGGCGTTGTACAATACCACCAATGAAGTA    333

LA_pLAH_nrsB_    551 TGCTACAAAGTGCTCAGGGATACTGGACGGATTGTCCTCCTTAACCTCAA    600
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual           334 TGCTACAAAGTGCTCAGGGATACTGGACGGATTGTCCTCCTTAACCTCAA    383

LA_pLAH_nrsB_    601 ATCTACGTGGATAGACATGATTGAAGGTTTCATGGAGGAAGCAAAATGGT    650
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual           384 ATCTACGTGGATAGACATGATTGAAGGTTTCATGGAGGAAGCAAAATGGT    433

LA_pLAH_nrsB_    651 TCAATGGTGGAAGTGCACCAAAATTGGAAGAGTATATAGAGAATGGAGTG    700
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
visual          434 TCAATGGTGGAAGTGCACCAAAATTGGAAGAGTATATAGAGAATGGAGTG   483

LA_pLAH_nrsB_   701 TCCACGGCAGGAGCATACATGGCTTTTGCACACATCTTCTTTCTCATAGG   750
                    |||||||||||||||||||||||||||||||||||||||||||||||||||
visual          484 TCCACGGCAGGAGCATACATGGCTTTTGCACACATCTTCTTTCTCATAGG   533

LA_pLAH_nrsB_   751 AGAAGGTGTTACACACCAAAATTCCCAACTCTTCACCCAAAAACCCTACC   800
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual          534 AGAAGGTGTTACACACCAAAATTCCCAACTCTTCACCCAAAAACCCTACC   583

LA_pLAH_nrsB_   801 CCAAGGTCTTCTCCGCCGCCGGCCGCATTCTTCGCCTCTGGGATGATCTC   850
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual          584 CCAAGGTCTTCTCCGCCGCCGGCCGCATTCTTCGCCTCTGGGATGATCTC   633

LA_pLAH_nrsB_   851 GGAACCGCCAAGGAAGAGCAAGAGCGAGGAGATCTGGCTTCGTGCGTGCA   900
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual          634 GGAACCGCCAAGGAAGAGCAAGAGCGAGGAGATCTGGCTTCGTGCGTGCA   683

LA_pLAH_nrsB_   901 GTTATTTATGAAAGAGAAGTCGTTGACGGAAGAGGAGGCAAGAAGTCGCA   950
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual          684 GTTATTTATGAAAGAGAAGTCGTTGACGGAAGAGGAGGCAAGAAGTCGCA   733

LA_pLAH_nrsB_   951 TTTTGGAAGAGATAAAAGGATTATGGAGGGATCTGAATGGGGAACTGGTC  1000
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual          734 TTTTGGAAGAGATAAAAGGATTATGGAGGGATCTGAATGGGGAACTGGTC   783

LA_pLAH_nrsB_  1001 TACAACAAGAATTTGCCGTTATCCATAATCAAAGTCGCACTTAACATGGC  1050
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual          784 TACAACAAGAATTTGCCGTTATCCATAATCAAAGTCGCACTTAACATGGC   833

LA_pLAH_nrsB_  1051 GAGAGCTTCTCAAGTTGTGTACAAGCACGATCAAGACACTTATTTTTCAA  1100
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual          834 GAGAGCTTCTCAAGTTGTGTACAAGCACGATCAAGACACTTATTTTTCAA   883

LA_pLAH_nrsB_  1101 GCGTAGACAATTATGTGGATGCCCTCTTCTTCACTCAATATCCCTATGAT  1150
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
visual          884 GCGTAGACAATTATGTGGATGCCCTCTTCTTCACTCAATATCCCTATGAT   933

LA_pLAH_nrsB_  1151 GTGCCCGATTATGCCTAA*GG-TCCCCGACCTGCAGGGGGGACGTT   1194
                    |||||||||||||||||||| |  ||.||
visual          934 GTGCCCGATTATGCCTAAGGAT--CCTAC----------------    960
```

Appendix 14: Full sequence and map of the pJ811 plasmid including the alkL gene (red)

Sequence

```
   1 TTAATCTTTC TGCGAATTGA GATGACGCCA CTGGCTGGGC GTCATCCCGG TTTCCCGGGT AAACACCACC
  71 GAAAAATAGT TACTATCTTC AAAGCCACAT TCGGTCGAAA TATCACTGAT TAACAGGCGG CTATGCTGGA
 141 GAAGATATTG CGCATGACAC ACTCTGACCT GTCGCAGATA TTGATTGATG GTCATTCCAG TCTGCTGGCG
 211 AAATTGCTGA CGCAAAACGC GCTCACTGCA CGATGCCTCA TCACAAAATT TATCCAGCGC AAAGGGACTT
 281 TTCAGGCTAG CCGCCAGCCG GGTAATCAGC TTATCCAGCA ACGTTTCGCT GGATGTTGGC GGCAACGAAT
 351 CACTGGTGTA ACGATGGCGA TTCAGCAACA TCACCAACTG CCCGAACAGC AACTCAGCCA TTTCGTTAGC
 421 AAACGGCACA TGCTGACTAC TTTCATGCTC AAGCTGACCA ATAACCTGCC GCGCCTGCGC CATCCCCATG
 491 CTACCTAAGC GCCAGTGTGG TTGCCCTGCG CTGGCGTTAA ATCCCGGAAT CGCCCCCTGC CAGTCAAGAT
 561 TCAGCTTCAG ACGCTCCGGG CAATAAATAA TATTCTGCAA AACCAGATCG TTAACGGAAG CGTAGGAGTG
 631 TTTATCATCA GCATGAATGT AAAAGAGATC GCCACGGGTA ATGCGATAAG GGCGATCGTT GAGTACATGC
 701 AGGCCATTAC CGCGCCAGAC AATCACCAGC TCACAAAAAT CATGTGTATG TTCAGCAAAG ACATCTTGCG
 771 GATAACGGTC AGCCACAGCG ACTGCCTGCT GGTCGCTGGC AAAAAAATCA TCTTTGAGAA GTTTTAACTG
 841 ATGCGCCACC GTGGCTACCT CGGCCAGAGA ACGAAGTTGA TTATTCGCAA TATGGCGTAC AAATACGTTG
 911 AGAAGATTCG CGTTATTGCA GAAAGCCATC CCGTCCCTGG CGAATATCAC GCGGTGACCA GTTAAACTCT
 981 CGGCGAAAAA GCGTCGAAAA GTGGTTACTG TCGCTGAATC CACAGCGATA GGCGATGTCA GTAACGCTGG
1051 CCTCGCTGTG GCGTAGCAGA TGTCGGGCTT TCATCAGTCG CAGGCGGTTC AGGTATCGCT GAGGCGTCAG
1121 TCCCGTTTGC TGCTTAAGCT GCCGATGTAG CGTACGCAGT GAAAGAGAAA ATTGATCCGC CACGGCATCC
1191 CAATTCACCT CATCGGCAAA ATGGTCCTCC AGCCAGGCCA GAAGCAAGTT GAGACGTGAT GCGCTGTTTT
1261 CCAGGTTCTC CTGCAAACTG CTTTTACGCA GCAAGAGCAG TAATTGCATA AACAAGATCT CGCGACTGGC
1331 GGTCGAGGGT AAATCATTTT CCCCTTCCTG CTGTTCCATC TGTGCAACCA GCTGTCGCAC CTGCTGCAAT
1401 ACGCTGTGGT TAACGCGCCA GTGAGACGGA TACTGCCCAT CCAGCTCTTG TGGCAGCAAC TGATTCAGCC
1471 CGGCGAGAAA CTGAAATCGA TCCGGCGAGC GATACAGCAC ATTGGTCAGA CACAGATTAT CGGTATGTTC
1541 ATACAGATGC CGATCATGAT CGCGTACGAA ACAGACCGTG CCACCGGTGA TGGTATAGGG CTGCCCATTA
1611 AACACATGAA TACCCGTGCC ATGTTCGACA ATCACAATTT CATGAAAATC ATGATGATGT TCAGGAAAAT
1681 CCGCCTGCGG GAGCCGGGGT TCTATCGCCA CGGACGCGTT ACCAGACGGA AAAAATCCA CACTATGTAA
1751 TACGGTCATA CTGGCCTCCT GATGTCGTCA ACACGGCGAA ATAGTAATCA CGAGGTCAGG TTCTTACCTT
1821 AAATTTTCGA CGGAAAACCA CGTAAAAAAC GTCGATTTTT CAAGATACAG CGTGAATTTT CAGGAAATGC
1891 GGTGAGCATC ACATCACCAC AATTCAGCAA ATTGTGAACA TCATCACGTT CATCTTTCCC TGGTTGCCAA
1961 TGGCCCATTT TCCTGTCAGT AACGAGAAGG TCGCGAATTC AGGCGCTTTT TAGACTGGTC GTAATGAAAT
2031 TCTTTTGTCG ACACTAGTCA TAGGAGGTAA AACATATGAG TTTTTCTAAT TATAAAGTAA TCGCGATGCC
2101 GGTGTTGGTT GCTAATTTTG TTTTGGGGGC GGCCACTGCA TGGGCGAATG AAAATTATCC GGCGAAATCT
2171 GCTGGCTATA ATCAGGGTGA CTGGGTCGCC AGCTTCAATT TTTCTAAGGT CTATGTGGGT GAGGAGCTTG
2241 GCGATCTAAA TGTTGGAGGG GGGGCTTTGC CAAATGCTGA TGTAAGTATT GGTAATGATA CAACACTTAC
2311 GTTTGATATC GCCTATTTTG TTAGCTCAAA TATAGCGGTG GATTTTTTTG TTGGGGTGCC AGCTAGGGCT
2381 AAATTTCAAG GTGAGAAATC AATCTCCTCG CTGGGAAGAG TCAGTGAAGT TGATTACGGC CCTGCAATTC
2451 TTTCGCTTCA ATATCATTAC GATAGCTTTG AGCGACTTTA TCCTTATGTT GGGGTTGGTG TTGGTCGGGT
2521 GCTATTTTTT GATAAAACCG ACGGTGCTTT GAGTTCGTTT GATATTAAGG ATAAATGGGC GCCTGCTTTT
2591 CAGGTTGGCC TTAGATATGA CCTTGGTAAC TCATGGATGC TAAATTCAGA TGTGCGTTAT ATTCCTTTCA
2661 AAACGGACGT CACAGGTACT CTTGGCCCGG TTCCTGTTTC TACTAAAATT GAGGTTGATC CTTTCATTCT
2731 CAGTCTTGGT GCGTCGTATG TTTTCCATCA TCATCATCAT CATTAATAAG CTAGCGAATT CAAGCTTGAT
2801 ATCATTCAGG ACGAGCCTCA GACTCCAGCG TAACTGGACT GAAAACAAAC TAAAGCGCCC TTGTGGCGCT
2871 TTAGTTTTTC TAGAGGATCC GTCGACCCCC TCATCCGAAA GGGCGTATTG GTACCGAGCT CGAATTCGTA
2941 ATCATGTCAT AGCTGTTTCC TGTGTGAAAT TGTTATCCGC TCACAATTCC ACACAACATA CGAGCCGGAA
3011 GCATAAAGTG TAAAGCCTGG GGTGCCTAAT GAGTGAGCTA ACTCACATTA ATTGCGTTGC GCTCACTGCC
3081 CGCTTTCCAG TCGGGAAACC TGTCGTGCCA GCTGCATTAA TGAATCGGCC AACGCGCGGG GAGAGGCGGT
3151 TTGCGTATTG GGCGCTCTTC CGCTTCCTCG CTCACTGACT CGCTGCGCTC GGTCGTTCGG CTGCGGCGAG
3221 CGGTATCAGC TCACTCAAAG GCGGTAATAC GGTTATCCAC AGAATCAGGG GATAACGCAG GAAAGAACAT
3291 GTGAGCAAAA GGCCAGCAAA AGGCCAGGAA CCGTAAAAAG GCCGCGTTGC TGGCGTTTTT CCATAGGCTC
3361 CGCCCCCCTG ACGAGCATCA CAAAAATCGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA GGACTATAAA
3431 GATACCAGGC GTTTCCCCCT GGAAGCTCCC TCGTGCGCTC TCCTGTTCCG ACCCTGCCGC TTACCGGATA
3501 CCTGTCCGCC TTTCTCCCTT CGGGAAGCGT GGCGCTTTCT CATAGCTCAC GCTGTAGGTA TCTCAGTTCG
3571 GTGTAGGTCG TTCGCTCCAA GCTGGGCTGT GTGCACGAAC CCCCCGTTCA GCCCGACCGC TGCGCCTTAT
3641 CCGGTAACTA TCGTCTTGAG TCCAACCCGG TAAGACACGA CTTATCGCCA CTGGCAGCAG CCACTGGTAA
3711 CAGGATTAGC AGAGCGAGGT ATGTAGGCGG TGCTACAGAG TTCTTGAAGT GGTGGCCTAA CTACGGCTAC
3781 ACTAGAAGGA CAGTATTTGG TATCTGCGCT CTGCTGAAGC CAGTTACCTT CGGAAAAAGA GTTGGTAGCT
3851 CTTGATCCGG CAAACAAACC ACCGCTGGTA GCGGTGGTTT TTTTGTTTGC AAGCAGCAGA TTACGCGCAG
3921 AAAAAAAGGA TCTCAAGAAG ATCCTTTGAT CTTTTCTACG GGGTCTGACG CTCAGTGGAA CGAAAACTCA
3991 CGTTAAGGGA TTTTGGTCAT GAGATTATCA AAAAGGATCT TCACCTAGAT CCTTTTAAAT TAAAAATGAA
4061 GTTTTAAATC AATCTAAAGT ATATATGAGT AAACTTGGTC TGACAGTTAG AAAAACTCAT CGAGCATCAA
4131 ATGAAACTGC AATTTATTCA TATCAGGATT ATCAATACCA TATTTTTGAA AAAGCCGTTT CTGTAATGAA
4201 GGAGAAAACT CACCGAGGCA GTTCCATAGG ATGGCAAGAT CCTGGTATCG GTCTGCGATT CCGACTCGTC
4271 CAACATCAAT ACAACCTATT AATTTCCCCT CGTCAAAAAT AAGGTTATCA AGTGAGAAAT CACCATGAGT
4341 GACGACTGAA TCCGGTGAGA ATGGCAAAAG TTTATGCATT TCTTTCCAGA CTTGTTCAAC AGGCCAGCCA
4411 TTACGCTCGT CATCAAAATC ACTCGCATCA ACCAAACCGT TATTCATTCG TGATTGCGCC TGAGCGAGAC
4481 GAAATACGCG ATCGCTGTTA AAAGGACAAT TACAAACAGG AATCGAATGC AACCGGCGCA GGAACACTGC
4551 CAGCGCATCA ACAATATTTT CACCTGAATC AGGATATTCT TCTAATACCT GGAATGCTGT TTTCCCGGGG
4621 ATCGCAGTGG TGAGTAACCA TGCATCATCA GGAGTACGGA TAAAATGCTT GATGGTCGGA AGAGGCATAA
4691 ATTCCGTCAG CCAGTTTAGT CTGACCATCT CATCTGTAAC ATCATTGGCA ACGCTACCTT TGCCATGTTT
4761 CAGAAACAAC TCTGGCGCAT CGGGCTTCCC ATACAATCGA TAGATTGTCG CACCTGATTG CCCGACATTA
```



Plasmid sequence and construct were provided from the supplier “gene art”. http://www.lifetechnologies.com/om/en/home/life-science/cloning/gene-synthesis/geneart-gene-synthesis.html

Appendix 15: Sequence results of the *alkL* gene with an HA tag

Key: *alkL* gene: (black), HA tag:(blue), Stop codon: (red)

```
Visual            1 ------Tatgagtttttctaattataaagtaatcgcgatgccggtgttgg     44
                        ||||||||||||||||||||||||||||||||||||||||||||||||||
Seq             451 AGGTCATATGAGTTTTTCTAATTATAAAGTAATCGCGATGCCGGTGTTGG    500

Visual           45 ttgctaattttgttttgggggcggccactgcatgggcgaatgaaaattat     94
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq             501 TTGCTAATTTTGTTTTGGGGGCGGCCACTGCATGGGCGAATGAAAATTAT    550

Visual           95 ccggcgaaatctgctggctataatcagggtgactgggtcgccagcttcaa    144
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq             551 CCGGCGAAATCTGCTGGCTATAATCAGGGTGACTGGGTCGCCAGCTTCAA    600

Visual          145 tttttctaaggtctatgtgggtgaggagcttggcgatctaaatgttggag    194
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq             601 TTTTTCTAAGGTCTATGTGGGTGAGGAGCTTGGCGATCTAAATGTTGGAG    650

Visual          195 gggggggctttgccaaatgctgatgtaagtattggtaatgatacaacactt    244
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq             651 GGGGGGCTTTGCCAAATGCTGATGTAAGTATTGGTAATGATACAACACTT    700

Visual          245 acgtttgatatcgcctattttgttagctcaaatatagcggtggatttttt    294
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq             701 ACGTTTGATATCGCCTATTTTGTTAGCTCAAATATAGCGGTGGATTTTTT    750

Visual          295 tgttggggtgccagctagggctaaatttcaaggtgagaaatcaatctcct    344
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq             751 TGTTGGGGTGCCAGCTAGGGCTAAATTTCAAGGTGAGAAATCAATCTCCT    800

Visual          345 cgctgggaagagtcagtgaagttgattacggccctgcaattctttcgctt    394
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq             801 CGCTGGGAAGAGTCAGTGAAGTTGATTACGGCCCTGCAATTCTTTCGCTT    850

Visual          395 caatatcattacgatagctttgagcgactttatccttatgttggggttgg    444
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq             851 CAATATCATTACGATAGCTTTGAGCGACTTTATCCTTATGTTGGGGTTGG    900

Visual          445 tgttggtcgggtgctatttttttgataaaaccgacggtgctttgagttcgt    494
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq             901 TGTTGGTCGGGTGCTATTTTTTGATAAAACCGACGGTGCTTTGAGTTCGT    950

Visual          495 ttgatattaaggataaatgggcgcctgcttttcaggttggccttagatat    544
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq             951 TTGATATTAAGGATAAATGGGCGCCTGCTTTTCAGGTTGGCCTTAGATAT   1000

Visual          545 gaccttggtaactcatggatgctaaattcagatgtgcgttatattccttt    594
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq            1001 GACCTTGGTAACTCATGGATGCTAAATTCAGATGTGCGTTATATTCCTTT   1050

Visual          595 caaaacggacgtcacaggtactcttggcccggttcctgtttctactaaaa    644
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq            1051 CAAAACGGACGTCACAGGTACTCTTGGCCCGGTTCCTGTTTCTACTAAAA   1100

Visual          645 ttgaggttgatcctttcattctcagtcttggtgcgtcgtatgttttcTAT    694
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq            1101 TTGAGGTTGATCCTTTCATTCTCAGTCTTGGTGCGTCGTATGTTTTCTAT   1150

Visual          695 CCCTATGATGTGCCCGATTATGCCTAA*GGAT--CCTAC-----------    730
                    |||||||||||||||||||||||||||| |  ||.||
Seq            1151 CCCTATGATGTGCCCGATTATGCCTAA*GG-TCCCCGACTGCATGGGGGGG   1199
```

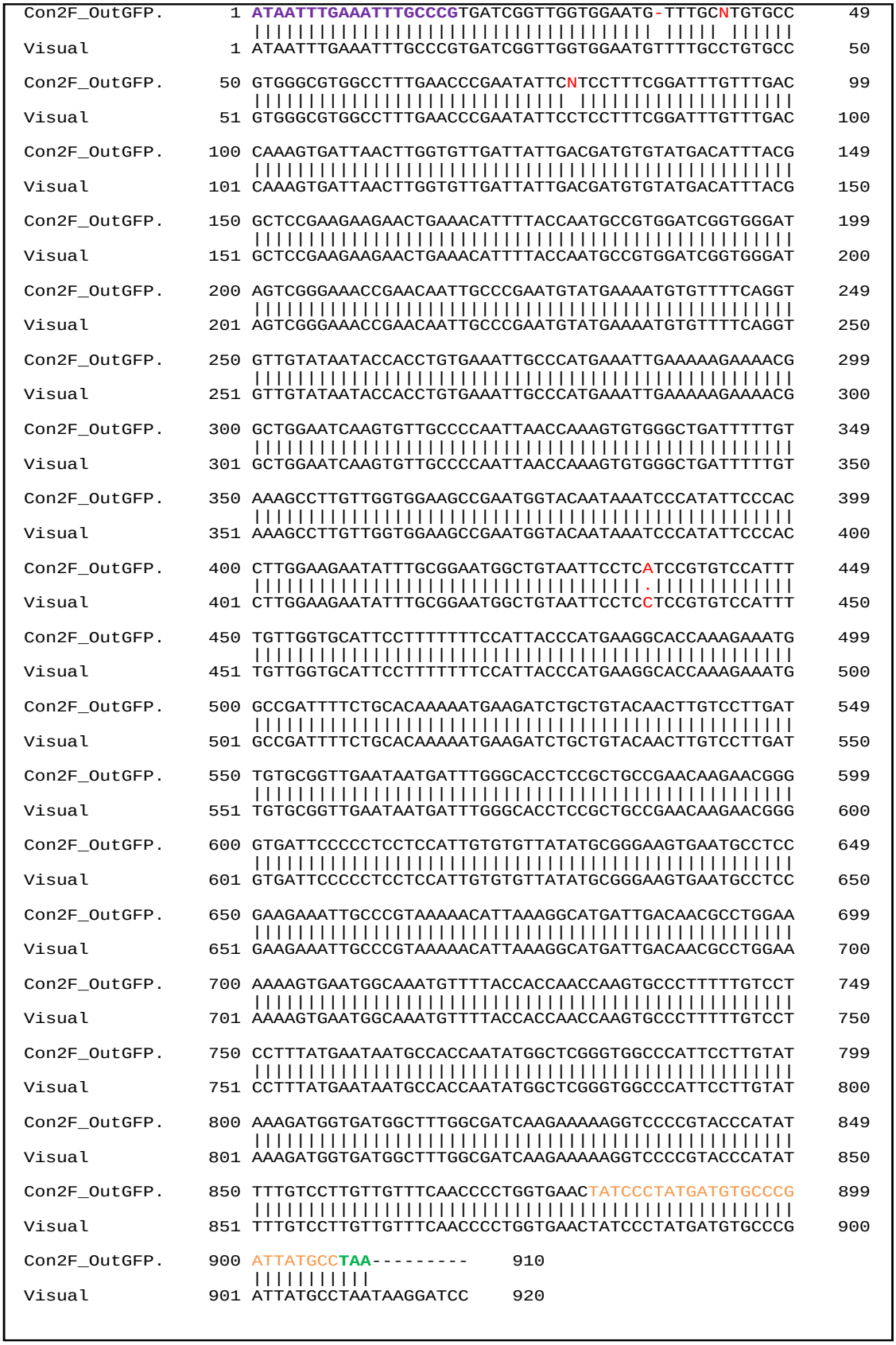
Appendix 16: List of websites used in this work.

figure #	website (URL)	access date
1.1.A	http://www.ucmp.berkeley.edu/bacteria/cyanolh.html	02/03/2014
1.1.B	http://ashonko.blogspot.com/2010/11/nostoc.html	02/03/2014
1.1.C	http://eol.org/data_objects/2086619	02/03/2014
1.1.D	https://microbewiki.kenyon.edu/index.php/Merismopedia	02/03/2014
1.1.E	http://schwarz10.de/tribal/urban_gardening/spirulina/index-en.html	02/03/2014
1.1.F	http://protist.i.hosei.ac.jp/pdb/Images/Prokaryotes/Chroococcaceae/Synechocystis/index.html	02/03/2014
1.2.B	http://palaeos.com/eukarya/glossary/glossaryP.htm	02/03/2014
1.2.C	http://www.jochemnet.de/fiu/bot4404/BOT4404_12.html	02/03/2014
1.3.A	http://cmore.soest.hawaii.edu/microscopy/album/symbionts/	02/03/2014
1.3.B	http://www.soils.umn.edu/academics/classes/soil3612/Symbiotic_Nitrogen_Fixation/Other.htm	02/03/2014
1.3.C	http://schuessler.userweb.mwn.de/geosiphon/geosiphon_home.html	02/03/2014
1.5	http://parts.igem.org/Help:Plasmid_backbones/Glossary	02/03/2014
1.7.A	http://academic.brooklyn.cuny.edu/biology/bio4fv/page/molecular%20biology/mutation-amestest.html	02/03/2014
1.8.B	http://www.motifolio.com/6111153.html	02/03/2014
1.10.B	http://www.emsl.pnl.gov/news/viewArticle.jsp?articleId=167	02/03/2014
1.15.A1	http://www.sigmaaldrich.com/catalog/product/sigma/i9516?lang=en&region=OM	03/03/2014
1.15.A2	http://en.wikipedia.org/wiki/File:1-Butanol_skeletal.svg	03/03/2014
1.15.A3	http://www.sigmaaldrich.com/catalog/product/aldrich/110949?lang=en&region=OM	03/03/2014
1.15.A4	http://en.wikipedia.org/wiki/Isobutanol	03/03/2014
1.15.B1	http://en.wikipedia.org/wiki/File:Limonene-2D-skeletal.svg	03/03/2014
1.15.B2	http://en.wikipedia.org/wiki/Farnesene	03/03/2014
1.15.C1	http://www.google.com/patents/US20120137574	03/03/2014
1.15.C2	http://www.google.com/patents/US20120137574	03/03/2014
1.17.A	http://biofuels.asu.edu/biomaterials.shtml	02/03/2014
1.17.B	http://www.me-newswire.net/news/exxonmobil-and-synthetic -genomics-inc-advance-algae-biofuels-program-with-new-greenhouse/en	02/03/2014
1.17.C	http://www.indiamart.com/jbnenterprises/algae-reactors.html	02/03/2014
1.17.D	http://www.powerplantccs.com/ccs/cap/fut/alg/algae_cultivation_systems.html	02/03/2014
1.17.E	http://inhabitat.com/new-submission-30/	02/03/2014

Appendix 17: Sequence analysis of α–farnesene synthase in the pLAH.A2 and pLAH.nrsB plasmids.

Sequence analysis of pLAH.A2.FS plasmid

```
Con2_F_HA_nrs    101 TGAGGTCATATGGAATTTCGGGTGCATTTGCATGCCGATCATGAACAAAA    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||
Visual             1 ------CATATGGAATTTCGGGTGCATTTGCATGCCGATCATGAACAAAA     44
Con2_F_HA_nrs    151 AATTCTGCAAAACCAGATGAAACCCGAACCCGAAGCCTCCTATCTGATTA    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Visual            45 AATTCTGCAAAACCAGATGAAACCCGAACCCGAAGCCTCCTATCTGATTA     94
Con2_F_HA_nrs    201 ATCAACGGCGTTCCGCCAATTACAAACCCAATATTTGGAAAAACGACTTT    250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Visual            95 ATCAACGGCGTTCCGCCAATTACAAACCCAATATTTGGAAAAACGACTTT    144
Con2_F_HA_nrs    251 CTGGACCAATCCTTGATTTCCAAATACGACGAAGATCAATATCGGAAACT    300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           145 CTGGACCAATCCTTGATTTCCAAATACGACGAAGATCAATATCGGAAACT    194
Con2_F_HA_nrs    301 GTCCGAAAAATTGATTGAAGAAGTGAAAATTTACATTAGTGCCGAAACCA    350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           195 GTCCGAAAAATTGATTGAAGAAGTGAAAATTTACATTAGTGCCGAAACCA    244
Con2_F_HA_nrs    351 AAGATTTGGTGGCCAAATTTGAACTGATTGATTCCGTGCGGAAATTGGGC    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           245 AAGATTTGGTGGCCAAATTTGAACTGATTGATTCCGTGCGGAAATTGGGC    294
Con2_F_HA_nrs    401 TTGGCCAATCATTTTGAAAAAGAAATTAAAGAAGCCTTGGATTCCATTGC    450
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           295 TTGGCCAATCATTTTGAAAAAGAAATTAAAGAAGCCTTGGATTCCATTGC    344
Con2_F_HA_nrs    451 CGCTATTGAATCCGATAATTTGGGCACCCGTGATGATTTGTATGGCACCG    500
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           345 CGCTATTGAATCCGATAATTTGGGCACCCGTGATGATTTGTATGGCACCG    394
Con2_F_HA_nrs    501 CCTTGCATTTTAAAATTTTGCGGCAACATGGCTACAAAGTGTCCCAAGAT    550
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           395 CCTTGCATTTTAAAATTTTGCGGCAACATGGCTACAAAGTGTCCCAAGAT    444
Con2_F_HA_nrs    551 ATTTTTGGTCGGTTTATGGATGAAAAAGGCACCCTGGAAAATCATCATTT    600
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           445 ATTTTTGGTCGGTTTATGGATGAAAAAGGCACCCTGGAAAATCATCATTT    494
Con2_F_HA_nrs    601 TGCCCATTTGAAAGGCATGTTGGAATTGTTTGAAGCCTCCAATTTGGGCT    650
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           495 TGCCCATTTGAAAGGCATGTTGGAATTGTTTGAAGCCTCCAATTTGGGCT    544
Con2_F_HA_nrs    651 TTGAAGGCGAAGATATTTTGGATGAAGCCAAAGCCTCCTTGACCTTGGCC    700
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           545 TTGAAGGCGAAGATATTTTGGATGAAGCCAAAGCCTCCTTGACCTTGGCC    594
Con2_F_HA_nrs    701 TTGAGGGATAGTGGCCATATTTGTTATCCCGATTCCAATTTGAGTCGGGA    750
                     |||.||||||||||||||||||||||||||||||||||||||||||||||
Visual           595 TTGCGGGATAGTGGCCATATTTGTTATCCCGATTCCAATTTGAGTCGGGA    644
Con2_F_HA_nrs    751 TGTGATTCATTCCTTGGAATTGCCCTCCCATCGGCGTGTGCAATGGTTTG    800
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           645 TGTGATTCATTCCTTGGAATTGCCCTCCCATCGGCGTGTGCAATGGTTTG    694
Con2_F_HA_nrs    801 ATGTGAAATGGCAAATTAATGCCTATGAAAAAGATATTTGTCGGGTGAAT    850
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           695 ATGTGAAATGGCAAATTAATGCCTATGAAAAAGATATTTGTCGGGTGAAT    744
Con2_F_HA_nrs    851 GCCACCTTGTTGGAATTGGCCAAACTGAATTTTAACATGGTGCAAGCCCA    900
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           745 GCCACCTTGTTGGAATTGGCCAAACTGAATTTTAACATGGTGCAAGCCCA    794
Con2_F_HA_nrs    901 ACTGCAAAAAGATTTGCGGGAAGCCTCC-AATGGTGGGCCAATTTGGGCA    949
                     |||||||||||||||||||||||||||| ||||||||||||||||||||
Visual           795 ACTGCAAAAAGATTTGCGGGAAGCCTCCAAATGGTGGGCCAATTTGGGCA    844
Con2_F_HA_nrs    950 TTGCCGATAATTTGAAATTTGCCCG
                     |||||||||||||||||||||||||
Visual           845 TTGCCGATAATTTGAAATTTGCCCG
```

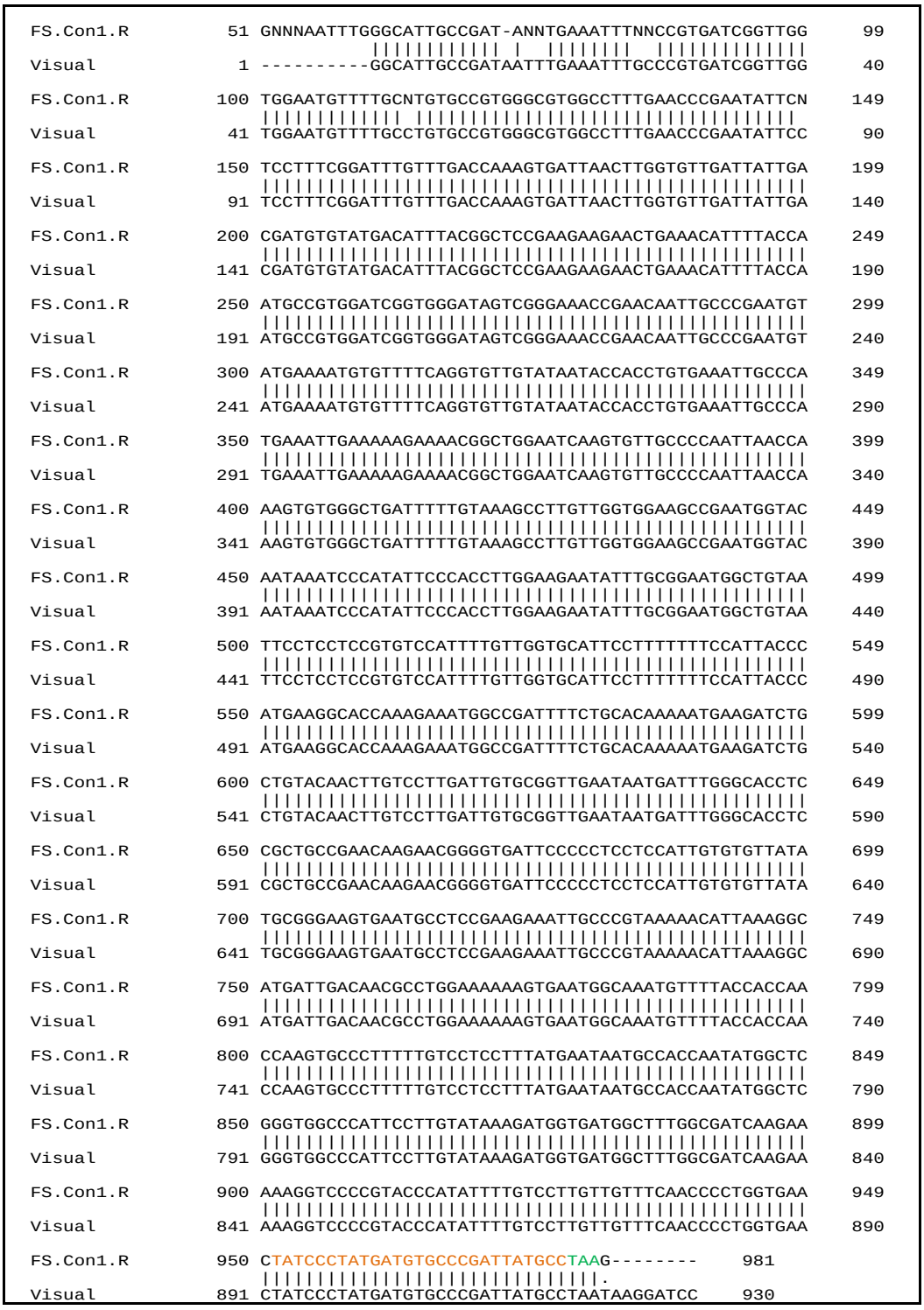


The two mutations in the pLAH.nrsB.FS plasmid were silent and did not affect the amino acid sequence.

Sequence analysis of pLAH.A2.FS plasmid

```
Con1.FS.HA       201 GAATTATAACCATATGGAATTTCGGGTGCATTTGCATGCCGATCATGAAC    250
                     ||||||||||||||||||||||||||||||||||||||||||||
Visual             1 ----------CATATGGAATTTCGGGTGCATTTGCATGCCGATCATGAAC     40

Con1.FS.HA       251 AAAAAATTCTGCAAAACCAGATGAAACCCGAACCCGAAGCCTCCTATCTG    300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Visual            41 AAAAAATTCTGCAAAACCAGATGAAACCCGAACCCGAAGCCTCCTATCTG     90

Con1.FS.HA       301 ATTAATCAACGGCGTTCCGCCAATTACAAACCCAATATTTGGAAAAACGA    350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Visual            91 ATTAATCAACGGCGTTCCGCCAATTACAAACCCAATATTTGGAAAAACGA    140

Con1.FS.HA       351 CTTTCTGGACCAATCCTTGATTTCCAAATACGACGAAGATCAATATCGGA    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           141 CTTTCTGGACCAATCCTTGATTTCCAAATACGACGAAGATCAATATCGGA    190

Con1.FS.HA       401 AACTGTCCGAAAAATTGATTGAAGAAGTGAAAATTTACATTAGTGCCGAA    450
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           191 AACTGTCCGAAAAATTGATTGAAGAAGTGAAAATTTACATTAGTGCCGAA    240

Con1.FS.HA       451 ACCAAAGATTTGGTGGCCAAATTTGAACTGATTGATTCCGTGCGGAAATT    500
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           241 ACCAAAGATTTGGTGGCCAAATTTGAACTGATTGATTCCGTGCGGAAATT    290

Con1.FS.HA       501 GGGCTTGGCCAATCATTTTGAAAAAGAAATTAAAGAAGCCTTGGATTCCA    550
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           291 GGGCTTGGCCAATCATTTTGAAAAAGAAATTAAAGAAGCCTTGGATTCCA    340

Con1.FS.HA       551 TTGCCGCTATTGAATCCGATAATTTGGGCACCCGTGATGATTTGTATGGC    600
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           341 TTGCCGCTATTGAATCCGATAATTTGGGCACCCGTGATGATTTGTATGGC    390

Con1.FS.HA       601 ACCGCCTTGCATTTTAAAATTTTGCGGCAACATGGCTACAAAGTGTCCCA    650
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           391 ACCGCCTTGCATTTTAAAATTTTGCGGCAACATGGCTACAAAGTGTCCCA    440

Con1.FS.HA       651 AGATATTTTTGGTCGGTTTATGGATGAAAAAGGCACCCTGGAAAATCATC    700
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           441 AGATATTTTTGGTCGGTTTATGGATGAAAAAGGCACCCTGGAAAATCATC    490

Con1.FS.HA       701 ATTTTGCCCATTTGAAAGGCATGTTGGAATTGTTTGAAGCCTCCAATTTG    750
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           491 ATTTTGCCCATTTGAAAGGCATGTTGGAATTGTTTGAAGCCTCCAATTTG    540

Con1.FS.HA       751 GGCTTTGAAGGCGAAGATATTTTGGATGAAGCCAAAGCCTCCTTGACCTT    800
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           541 GGCTTTGAAGGCGAAGATATTTTGGATGAAGCCAAAGCCTCCTTGACCTT    590

Con1.FS.HA       801 GGCCTTGCGGGATAGTGGCCATATTTGTTATCCCGATTCCAATTTGAGTC    850
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           591 GGCCTTGCGGGATAGTGGCCATATTTGTTATCCCGATTCCAATTTGAGTC    640

Con1.FS.HA       851 GGGATGTGATTCATTCCTTGGAATTGCCCTCCCATCGGCGTGTGCAATGG    900
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           641 GGGATGTGATTCATTCCTTGGAATTGCCCTCCCATCGGCGTGTGCAATGG    690

Con1.FS.HA       901 TTTGATGTGAAATGGCAAATTAATGNCTATGAAAAANATATTTGTCGGGT    950
                     |||||||||||||||||||||||||| |||||||||| ||||||||||||||
Visual           691 TTTGATGTGAAATGGCAAATTAATGCCTATGAAAAAGATATTTGTCGGGT    740

Con1.FS.HA       951 GAATGCCACCTTGTTGGAATTGGCCAAACTGAATTTTAACATGGTGCAAG   1000
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Visual           741 GAATGCCACCTTGTTGGAATTGGCCAAACTGAATTTTAACATGGTGCAAG    790

Con1.FS.HA      1001 CCCAACTGNNNNNN-----------------------------------   1014
                     ||||||||
Visual           791 CCCAACTGCAAAAAGATTTGCGGGAAGCCTCCAAATGGTGGGCCAATTTG    840
```

(The two sequences do not overlap since 42 base pairs have not been sequenced due to bad DNA signal at the end of the chromatogrames)



Key: *Nde*I site (blue), over-lapping sequences between forward and reverse DNA sequence reactions (purple), HA tag (orange), stop codon (green).

Publications